ADA 283580

THE SYSTEMS ENGINEERING APPROACH TO

MECHANICAL FAILURE PREVENTION

Proceedings of the 47th Meeting
of the
Mechanical Failures Prevention Group

Virginia Beach, Virginia
April 13-15, 1993

94-26595

Compiled by
Henry C. Pusey
and
Sallie C. Pusey

A Publication of the

Vibration Institute

A NOT-FOR-PROFIT CORPORATION

Special Notice

ii

TABLE OF CONTENTS

CONDITION BASED MAINTENANCE SYSTEMS ENGINEERING

PREFACE

The 47th Meeting of the Mechanical Failures Prevention Group(MFPG) was sponsored by the Office of Naval Research(ONR), Arlington, VA; the Naval Surface Warfare Center(NSWC), Annapolis, MD; the Naval Civil Engineering Laboratory(NCEL), Port Hueneme, CA; the U.S. Army Research Laboratory, Watertown, MA and the Vibration Institute. The conference was held April 13-15, 1993 at the Cavalier Hotel in Virginia Beach, Virginia. Meeting management, program coordination, and proceedings compilation were by the Vibration Institute. MFPG Council Chairman G. William Nickerson chaired the Opening Session. The Poster Session Coordinator and Session CoChairmen are identified on the title pages for each section in these proceedings. The MFPG Council and the MFPG Program Committee Members are listed separately.

We were especially pleased this year to have Captain Peter Child from the Canadian National Defence Headquarters as our Keynote Speaker. Captain Child's paper, along with three Opening Session papers (Speakers Hansen, Bently and Richardson) and a Plenary Paper (Speaker Pecht), are included in the FEATURED PAPERS section of these proceedings. Regrettably, three distinguished invited speakers presented excellent papers that are not included in the proceedings. Mr. Leonard S. Tedesco of the Ford Motor Company spoke on *Diagnostics of Automotive Electronic Systems* in the Opening Session. On the second day, Mr. Ernest J. Czyryca from the Naval Surface Warfare Center gave a Plenary Address on *Lessons Learned in Metallurgical Failure Analyses of Naval Machinery*. The final Plenary Lecture on *Durable High Performance Blading* was presented by Dr. Neville F. Rieger, President of Stress Technology, Incorporated.

The MFPG Technical Program also included three mini courses, an evening workshop and a final afternoon panel/workshop. The mini courses presented were as follows

1. *An Introduction to Wear of Engineering Materials*: Dr. Said Jahanmir, National Institute of Standards and Technology, Gaithersburg, MD.

2. *Assessing the Economic Value of Mechanical Failure Prevention*: Professor Wolter J. Fabrycky, Virginia Polytechnic Institute and State University, Blacksburg, VA.

3. *Signal Processing for Diagnostics*: Dr. C. James Li, Columbia University, New York, NY

A special working group was formed as the result of a recommendation made during a panel session at MFPG 46. At MFPG 47 the working group conducted a workshop on *The Business Case for Mechanical Failure Prevention*. Ms. Karen Krewer from the Office of the Chief of Naval Operations (formerly from NAVSEA) chaired the Workshop. She was assisted by working group members John Major from Newport News Shipbuilding and S. Nils Straatveit from General Dynamics/Electric Boat. The closing panel session on *Applications of Neural Networks in Mechanical Failure Prevention* was cochaired by Mr. Joel Milano, Naval Surface Warfare Center, Bethesda, MD and Mr. James W. Taylor, HSB Reliability Technologies, Inc., Arlington, VA. The members of the panel were Dr. C. James Li, Columbia University; Dr. James Lo, University of Maryland at Baltimore; Dr. Reginald G. Mitchiner, Virginia Polytechnic Institute; Mr. Richard Morris, Naval Surface Warfare Center; and Dr. Young Shin, Naval Postgraduate School.

The Mechanical Failures Prevention Group was organized in 1967 under the sponsorship of the Office of Naval Research. The MFPG was formed for the express purpose of stimulating and promoting voluntary cooperation among segments of the scientific and engineering communities in order to gain a better understanding of the processes of mechanical failures. The goals were to reduce the incidence of mechanical failures by improving design methodology, to devise methods of accurately predicting mechanical failures and to apply the increased knowledge of the field to the problems of our present technology. Through the activities of its Technical Committees the MFPG continues to act as a focal point for any technological developments that contribute to mechanical failure reduction or prevention. The purpose of the work of the Technical Committees is to

♦ Collect, analyze, and disseminate technical information on mechanical failures.

♦ Facilitate the transfer of technology from government to the private sector.

♦ Establish appropriate terminology, criteria and terms of reference.

♦ Critically examine the field of mechanical failures to determine needed areas of endeavor and make suitable recommendations.

♦ Provide advisory recommendations and technical expertise in the field.

♦ Encourage research and development directed toward both the prior identification and the reduction of mechanical failures.

♦ Maintain awareness of all significant work relevant to the identified interest areas.

♦ Stimulate interdisciplinary communication among those who can contribute technically, and provide a suitable forum for their direct discussions through meetings, conferences and symposia.

♦ Periodically review the state-of-the-art of mechanical failure technology; facilitate transition of new laboratory developments into hardware capable of alleviating operational problems.

♦ Identify areas in research and development where effort is disproportionate to promise and recommend action as deemed necessary.

Those interested in working on any of the Technical Committees should contact the appropriate committee chairman. The committees, along with the names and addresses of the chairmen, are included in the MFPG Council listing.

On behalf of Dr. Eshleman and the Vibration Institute, I want to thank our co-sponsors and the MFPG Council for their cooperation in organizing and conducting the 47th MFPG Meeting. We are exploring some exciting possibilities for the future and fully expect that our conferences will continue to provide an effective forum for those who have mechanical failure problems and those who are engaged in failure avoidance technology.

Henry C. Pusey
Executive Secretary

FEATURED PAPERS

Opening Session and Plenary

CANADIAN NAVAL MAINTENANCE

Captain(N) P. J. Child
Director of Ship Engineering
National Defence Headquarters, Ottawa, Canada, K1A OK2

Abstract: This paper will address the Canadian naval maintenance philosophy and the framework in which it exists, the monitoring and data systems in place and planned. The paper will conclude by addressing the expectations from the systems that are planned.

Maintenance framework: The fact that the Canadian Fleet must operate in the geopolitical environment of the 1990s is a given. The demise of the solid threat focus from the Cold War has brought about an emerging focus on sovereignty and a significant reduction in the focus on anti-submarine warfare. Perhaps of more significance is the public perception of a "peace dividend".

The peace dividend has placed significant pressures on the defence dollar, a dollar that was already under severe attack internally as the Department sought to increase the investment in new capital equipment and replace the old equipment which were suffering from a condition described by many as "rust out".

Canadian Navy Maintenance Philosophy: Throughout the 1970s, Preventive Maintenance activities continued to be time based. Considerable resources were expended to maintain and overhaul equipment whether its condition warranted it or not. However, as time passed, economy measures have steadily eroded both manpower and maintenance resources. At the same time, increasingly more complex systems and equipments were being fitted to ships, requiring more involved maintenance, repair and testing procedures. In the early 1980s, the drive to ensure that optimum benefit would be realized from the maintenance effort brought about a change of maintenance concept. Naval maintenance went from a time based philosophy to a reliability centred maintenance (RCM) and a Maintenance Requirements Analysis approach. The determination of the maintenance requirements and the application of the resources to satisfy these requirements were promulgated in a 1984 Maintenance Policy Statement.

The new maintenance concept, based on reliability centred maintenance, was predicated on the achievement of a balance between the resources available and the degree of operational availability

desired. The policy required the new concept to be applied to new ships and to all new equipments being acquired. Using analytical techniques it would be determined for each equipment and system whether preventive maintenance would be done at all; if so, whether it would be time based or condition based; and then, what maintenance work would be performed. Time-based preventive maintenance was retained where safety requirements dictate that every precaution must be taken to prevent failure, where continued availability of the system is operationally essential, and for systems which do not lend themselves to condition based maintenance.

Equipment Health Monitoring (EHM) techniques will be used in the assessment of equipment condition. The goal ascribed to EHM is the determination of the condition of equipments and systems in order to assist in maintenance decision making that will maximize the service life and availability. EHM for electronic equipment is largely accomplished through built-in test equipment, external test procedures and diagnostic routines. For mechanical equipment, performance indicators are obtained form analytical methods utilizing fluid analysis and a variety of instrumentation systems to measure appropriate performance parameters. For hull systems, EHM techniques consist of inspections, measurements and non-destructive tests.

Condition based monitoring systems and processes: The use of fluid analysis for EHM in the Canadian Navy can be traced back to the late 1960s when the concept was first introduced. Around 1970, the first fleet trials of spectrometric oil analysis were initiated. Between 1970 and 1978, the concept was developed into the Spectrometric Oil Analysis Program (SOAP). In the ensuing years, there was a somewhat haphazard approach to the development and application of Spectrometric Oil Analysis which resulted in an over-sized program with limited credibility in the eyes of the maintainers.

In 1985, a pilot program was initiated to avoid the deficiencies and limitations of SOAP. This program expanded the scope of the analysis to include coolant testing and was named the Oil and Coolant Condition Analysis Program (OCCAP). The program was adopted in 1991 and the key aspects of the approach are:

a. careful screening to ensure inclusion only of those equipments which, due to their design characteristics and critical nature and/or high costs, benefit from fluid monitoring;

b. fully automated information management with centralized database control and expert system technology applied to data interpretation; and

c. the contracting of the sample analysis to private industry.

OCCAP verifies that correct lubricants and coolants are being used, ensures that important fluid properties are maintained in service, and assists in assessing overall equipment condition. The previous

SOAP, which looked at wear metals only, neglected the importance of lubricant and coolant quality (viscosity, fuel dilution, etc.) to equipment performance. Note that the time from sample pick-up at the ship to availability of the OCCAP report generally does not exceed two working days in home port!

The OCCAP's relational database software consists of a series of tables which contain information pertaining to the system users, ships and land installations, sampled equipment, lubricants, coolants, and analysis results. The knowledge base contains several hundred "if-then-else" type rules which were developed after extensive knowledge engineering. Although the original knowledge base was purchased from IFS Corporation, rules are being added and deleted from the system by the Department on an on-going basis as OCCAP is refined.

The use of artificial intelligence provides precise and consistent recommendations that are based on a wider set of parameters than would be the case for a subjective human assessment. Knowledge and experience are also captured, making OCCAP less sensitive to the effects of a transient work force. This benefit was demonstrated when an interview of a West Coast technician resulted in the addition of approximately 25 new rules prior to his transfer.

A facet of diesel engine monitoring under OCCAP which is significant, although not directly quantifiable, is the impact on safety. OCCAP monitors the flash point of all diesel lubricant samples and there have been cases where samples have failed the flash point requirement of less than 190 deg C. This limit has been established by NDHQ as a safety minimum, and is not an operational limit in terms of machine performance. The value of detecting low flash point is indisputable given the serious consequences of crank case explosions.

Vibration analysis techniques are employed as a principle method for condition monitoring. As a routine maintenance procedure, Canadian naval ships conduct vibration surveys several times a year. Vibration monitoring blocks are fitted at predetermined locations on all rotating equipment to facilitate good repeatability of measurements.

When completed, the survey results are compared against fleet norms to determine the extent and nature of deterioration. Fleet vibration norms are maintained by summing and averaging fleet vibration survey records. Specific octave band fluctuations can then be traced to bearing wear or other rotational imbalances and hence decisively determine the appropriate corrective maintenance procedure.

In the past, the conduct of vibration surveys was time consuming and often involved extensive delays during the fleet norm comparison procedure. The current Canadian naval techniques have streamlined this process and allow an immediate fleet norm comparison.

The key instrument in the new procedure is a portable vibration logging device called the "DATA-TRAP". This commercial equipment has

5

undergone a series of military modifications to allow data
compatibility with existing fleet norms and to simplify operating
procedures. The DATA-TRAP is used to capture and store vibration
records following a pre-programmed survey route, while noting that
the monitored equipment is operating at specified load and speed
conditions. A route survey requires only that the operator place a
hard wired transducer, which employs a magnetic base, to the
appropriate vibration monit)ring block for approximately 30 seconds.
The survey data is then transferred to a standard personal computer
onboard the ship for processing. The ensuing computation compares
fleet norms against the current survey and immediately identifies
anomalies.

An Artificial Intelligence based "Expert System" is under
development at the Naval Engineering Test Establishment to further
enhance the DATA-TRAP post-processing diagnostic capabilities. This
system will include many equipment specific details and guide the
novice user through a series of more complex analyses aimed at
reducing maintenance down time.

As the effects of condition based maintenance are fully realized,
and historical time based maintenance practises curtailed,
systematic equipment health monitoring procedures will become
indispensable. In order to adapt to a Short Work Period rather than
a planned refit maintenance schedule, it will become the rule rather
than the exception to provide documented evidence that a maintenance
procedure is justified. Specific, rather than total overhauls, will
be conducted and additional information regarding the nature of a
equipment defect will be required. The Canadian Naval vibration
analysis program using the DATA-TRAP data logger has proved well
suited for this task.

Refits: In the early 1960s the fleet functioned with an operational
cycle of approximately 12 months followed by an assisted maintenance
period of about 4 months. During this period, virtually all
maintenance was performed on a periodic basis based on operating
hours or the calendar. In the late 1960s, it was appreciated that
there was considerable over maintenance and the periodicity was
moved to a 20 and then a 24 month cycle. The next step, taken in
the early 1970s was to implement a 1 year long baseline refit after
3 years of operations (a 4 year maintenance profile) with 4, three
week assisted-maintenance periods per year during the operational
cycle. In the early 1980s the maintenance profile was extended to 5
years as the driving underwater hull corrosion problems were solved.

The baseline refits returned virtually all equipments, particularly
the rotating machinery, to a known baseline, virtually as-new
condition every four years in order to provide confidence that the
vessel could operate for the fully operational cycle with as small a
risk of equipment failure as possible. We literally opened the
equipment to see why it was working so well and, on return to
service, we had frequently injected faults. The baseline refit
philosophy was driven by the fact that the fleet was aging with few

6

perceived opportunities in the foreseeable future for replacement. We had to make the equipments last as long as possible and, during the refits, we took the opportunity to install as many engineering changes as we could develop and afford in an attempt to have the fleet remain operationally viable. Typically, we would budget about 25% of the available person hours for the installation of change and the remainder for the refit including the rebuilding of machinery. The assessment of condition was not particularly important during this period and the question most frequently asked was "Can you guarantee that this piece of equipment will not fail during the next four years if it is not looked at?"

With the approval of the Canadian Patrol Frigate Project in 1983 and the knowledge of the retirement dates for the existing steam driven fleet, the Navy introduced two different refitting philosophies:

a. The first was "condition based refit" and was applied to vessels which were projected to have over six years of operations remaining. Under this philosophy, non-safety related equipments were refitted only if there was a condition deterioration as registered by oil analysis and vibration analysis to justify the requirement. In practice, even this philosophy was subverted quite easily in that the operating period was four years before the next significant refitting availability and any deterioration in operational capability from the "new" condition became the excuse by Commanding Officers and their engineers to have the equipment refitted. They were not accountable for the resources used by either the second or third line in support of their vessel. The degree of unnecessary expenditures caused by this process is not known but there is and has been the suspicion that we over-maintain our ships.

b. For a vessel entering the final planned refitting period we introduced a "care and custody" refit which was allocated about half the resources of the baseline refit and was intended to repair known failures only. This approach was more successful in limiting the amount of work performed but was not matched by any changes in the attitudes of the sea-going personnel who still endeavoured to have all systems functioning at all times. The maintenance load started to shift from third line to second line in the Ship Repair Units.

The mandated maintenance philosophy of the Canadian Patrol Frigate Project is essentially a phased maintenance approach with repair by replacement (RxR) and maintenance by exchange (MxE) during four, three week, short work periods per year and no refit planned until a modernization period after twelve years. There is a provision for extended work periods every four years to allow the docking work to be performed. The RxR and MxE philosophy is heavily condition based and the ship is equipped with "bite" and diagnostic equipment. The short work periods will be characterized by three types of work being undertaken: running repairs, progressive overhaul and the installation of change.

The dwindling size of the fleet in response to "peace breaking out all over", is directing a current look into the phased maintenance philosophy for the ships other than the new frigates. We are actively addressing the implications of converting our replenishment ships to phased maintenance driven by the fact that we will soon be operating but two of these vessels with two sides of the continent to address. The frequent work periods of this approach will allow the monitoring of deteriorating condition and the planning of repairs when condition deterioration has reached an unacceptable level. The data systems in support of phased maintenance are of particular importance.

Data systems: In the 1970's, an early attempt to automate some functions of maintenance management resulted in the Ships Maintenance Management Information System (SMMIS). SMMIS was and is still using manual input forms to update a central data base, which in turn is used to generate voluminous reports. It was originally developed for Tribal Class Destroyers. Other ship classes and submarines were incorporated into the system in subsequent years. Minesweepers, Auxiliary and Reserve Vessels have never been included in SMMIS.

SMMIS captures only first and second line repair facility data for preventive maintenance; corrective maintenance; equipment transfers; engineering changes; and miscellaneous maintenance actions. Third line (refit) maintenance information has never been captured by SMMIS.

SMMIS has suffered from neglect. As a result, the quality of data has deteriorated and SMMIS credibility is suspect. The major deficiencies of the naval maintenance management are:

a. SMMIS data is suspect and the time lag between maintenance events and reports is too great. Many of the inaccuracies came from the fact that the system did not offer any tangible benefit to the personnel who were charged with inputting the data;

b. naval maintenance management and reporting is assisted, and sometimes controlled, by numbers of separate computer-based information systems. While there are attempts to coordinate the functions of the Naval maintenance system, there is no clear path for data flow between the various systems nor do these systems provide all the tools required by NDHQ Staff, Command, shore maintenance units, or ship's staff; and

c. there are no automated shipboard facilities to collect maintenance data or provide ship staff with the up-to-date configuration.

The recognition of these deficiencies has resulted in a concept to go far beyond the narrow, imposed boundaries of SMMIS. The new system, the Naval Maintenance Management Information System (NAMMIS) is aimed at a broad spectrum of maintenance functions, designed to

8

assist both ship and shore-based engineering and maintenance personnel.

The development of NAMMIS will take several years. Prior to final delivery, all CPF (HALIFAX Class) and post-TRUMP IROQUIOIS Class ships will have been accepted into the Fleet. As we cannot afford to wait until then, it was decided that a progressive approach would be taken that would see the continued development of NAMMIS and at the same time:

a. confirm the requirements for an automated onboard maintenance capability during a trial project;

b. pending the trial's success, install similar hardware and software onboard HALIFAX and IROQUOIS Class ships; and

c. improve SMMIS by:

   (1) eliminating any duplicate or unwanted reports;

   (2) replacing the hand written input process to update the database by electronic entry;

   (3) developing a new menu-driven "extracto" process to make it easier and much quicker to produce ad hoc reports; and

   (4) correcting SMMIS' most salient problem (the quality and accuracy of the information reported) by setting up a Quality Control process.

In 1988, a two year trial onboard HMCS HURON was initiated. The shipboard installation consisted of a Local Area Network (LAN) composed of a file server and 10 personal computers and printers. A wide variety of created and off-the-shelf software were made available on this network. The final report concluded that the objectives were accomplished and it was recommended to install a similar system onboard most classes of ships. The Integrated Configuration and Engineering Maintenance Network (ICEMaN) project, as it became known, is targeted for installation starting in 1993. ICEMaN automates many of the administrative functions of ship maintenance, as given below:

a. **Maintenance Administration:** the collection and tracking of maintenance information;

b. **Preventive Maintenance:** the production of preventive maintenance schedules and lists;

c. **Short Work Period:** the establishment of priorities and the tracking of Repair Facility maintenance activities;

d. **Equipment Record Register:** the inventory of fitted shipboard equipment;

9

e. **Reliability, Availability, Maintainability:** the collection and tracking of Reliability, Availability, Maintainability information;

f. **Equipment Health Monitoring:** the scheduling of Equipment Health Monitoring tests as well as recording and analyzing the results;

g. **Unsatisfactory Condition Report:** the recording and tracking of Unsatisfactory Condition Report information;

h. **Supply:** the inventory of shipboard supplies as well as the recording of Supply Document information; and

j. **Other functions:** in addition various commercial software (e.g., Wordperfect, Draw Perfect, Lotus 1-2-3, and ORACLE) and other DND-developed applications are accessible from any work station on the shipboard network.

The Naval Maintenance Management Information System (NAMMIS) is a DND Capital project estimated at approximately $25M and scheduled for completion prior to the end of the century. NAMMIS will bring the various maintenance information/configuration systems under a common umbrella and provide access to DND information systems such as the supply system and the financial system. It is the logical continuation of the ongoing, long-term planning process for naval maintenance.

Conclusions: Our experience can lead to what I consider to be several significant conclusions:

a. condition based maintenance is not easy to implement and is perhaps not appropriate if the maintenance profile includes a refit periodicity of five years;

b. a design with redundant systems and equipments cannot be used to its cost effective end if the operators of the ships insist that they sail with all systems, both primary and secondary, functional at all times;

c. in order to effectively employ a condition based maintenance philosophy, one must be prepared to accept risk;

d. a system which provides no accountability to the ship's company for the amount of second and third line resources that their vessel consumes, will not encourage the cost effective employment of resources; and

e. the changing of the culture is much more difficult than the changing of the hardware.

Our expectations from RCM and CBM are not diminished. I am firmly convinced that we have traditionally over-maintained our ships in the interests of minimizing risks and this is not an appropriate

action given the costs and the environment in which we are now living. Although I applaud the advances that have occurred in the field of mechanical failures prevention, we must carefully look at the business case that can be created. I am left with the conviction that the technical problems that the Mechanical Failures Prevention Group is seeking to solve are, in reality, the easy part and that we should perhaps be dedicating more of our effort to promoting belief in the results.

UNIVERSITY ROLE IN MECHANICAL FAILURE PREVENTION

Robert J. Hansen
Applied Research Laboratory
The Pennsylvania State University
Post Office Box 30
State College, PA 16804

Stewart K. Kurtz
Materials Research Institute
The Pennsylvania State University
University Park, PA 16802

Abstract

United States universities presently make only a modest contribution to the development of improved methods of mechanical diagnostics and failure prevention. The reason is not a shortage of excellent instruction and research on individual topics important to the area (e.g., fracture mechanics, signal processing), but rather a failure to integrate these topics into multidisciplinary courses and research thrusts specifically addressing diagnostics and failure prevention. In the present paper, a research and education agenda for university-based activity pertaining to mechanical reliability and diagnostics for the rest of this decade is proposed. It focuses on (a) the development of prognostics capabilities which both identify failure precursors and accurately predict the remaining time to failure, (b) understanding of the dominant failure mechanisms associated with the new materials and materials processing methodologies that will be introduced for improved mechanical system component reliability, and (c) the formulation of undergraduate engineering curricula which prepare engineers to fully account for the needs and opportunities presented by reliability engineering and condition monitoring concepts in design and manufacturing.

Key Words: Diagnostics; prognostics; condition-based maintenance; reliability; wavelet transforms; nonlinear dynamics.

Introduction

Mechanical system reliability and maintainability have become areas of growing concern in recent years in both the civilian and military sectors because of their impact on human safety and system lifecycle cost. Such disasters as the breakup of a part of the fuselage of an Aloha passenger jet at cruising altitude points to the need for improved capability to predict safe life and anticipate incipient failures. By the same token, the aging infrastructure of roads and bridges present a potentially massive maintenance expenditure,

13

and the same remaining safe life question is important to prioritizing the order in which work should be accomplished and anticipating (and hopefully avoiding) disasters such as occurred in the collapse of an Interstate 95 bridge in southwestern Connecticut. Within both the defense and civilian sectors, the safety of helicopter pilots depends upon the ability to anticipate gearbox failures. At the same time, lifecycle cost considerations are best served by performing maintenance when needed, rather than in accordance with a conservative, worst case scenario maintenance schedule. Cost considerations have also motivated the Electric Power Research Institute to establish a Monitoring and Diagnostics Center at the Philadelphia Electric Company to serve the particular interests of electric utility companies.

Unfortunately, the importance society has come to associate with mechanical system reliability and maintainability is not yet reflected in the attention devoted to them in university engineering curricula and research projects. Probabilistic design life methods have been developed and are now in wide use in the electronics and aircraft sectors [3], but the typical undergraduate engineer is given little or no training in these methods and so is unacquainted with their utility in the design process. The same may be said of the awareness imparted of the design implications of machinery condition monitoring and condition-based maintenance. At the graduate level, many research projects are conducted on individual topics potentially relevant to mechanical failure prevention, such as fracture mechanics, signal analysis, and decision strategies. Unfortunately, almost none are undertaken as part of a multidisciplinary research thrust which has as its goal the development of new or improved reliability predictions or condition monitoring approaches. Moreover, this modest U.S. university involvement contrasts with that of the universities of some of our major technological competitors.

The purpose of the present paper is to suggest a reliability/maintainability research and instruction agenda for the U.S. university community for the rest of this decade. The recommendations offered for undergraduate instruction should be widely implemented in our judgment. The research agenda may best be accomplished by the establishment of one or more university-based centers of excellence addressing specific aspects of mechanical failure prediction.

U.S. University Role to Date

Reliability engineering at universities has had a long if uneven history. The early period, which dates from the first half of the twentieth century was devoted to the durability of important mechanical structures such as bridges, high rise buildings, and aircraft as well as the materials of which they were constructed. Emphasis was on mechanical fatigue and structural failure under various conditions of loading. By separate evaluation of the statistical distribution of externally applied stresses in a given application, as well as the statistical distribution of the strength for a given structural element, a design could be specified by minimizing the overlap regions between the maximum applied stresses and the minimum strength. Major emphasis was thus placed on safety margins, design guidelines, and guard bands, rather than the accurate prediction of the probability of failure at a given instant or hazard rate. The hazard rate $h(t)$ is the instantaneous probability of failure in the time internal t to $t + \Delta t$ for a device or non-repairable structure assuming it has survived up to time t.

14

$$h(t) = \frac{d \ln R(t)}{dt}$$

where R(t) is the reliability function or probability of survival. In a sense, it is the components "survival signature".

The second phase of reliability engineering originated in the requirements of the armed services for reliable electrical and electronic components and systems for use under wartime conditions. Early efforts at the Army-Navy facilities led to the JAN (Joint Army-Navy) specifications for vacuum tubes and electrical components used in communications receivers and transmitters, radars, bombsights, sonars, and the mass of new technology which was instrumental in helping the Allies achieve victory in World War II, often against numerically superior forces. A history of the development of electronic reliability can be found in Pollino [13]. The emphasis again was on extensive testing and use of an empirical approach to provide safety margins and design quicklimes. Minimal emphasis was placed on reliability modeling or attempting accurate predictions of the time to failure, or the specification of hazard rates. For this reason, electrical and electronic reliability was very slow in entering university curricula with few exceptions. One notable exception was the electric power industry where high reliability was essential to the delivery of electric power to customers. Power outages and brown-outs being both costly and creating a significant public hazard, were unacceptable. Thus, some university courses in power systems reliability and the related textbooks and scholarly activity can be traced back to this early phase.

We note that during these first two phases of reliability engineering, the scientific and scholarly study of the physical mechanisms by which failure occurred received limited attention. There were notable exceptions such as the seminal research of Shinozuka and Gumbel at Columbia and Freudenthal at George Washington University on fracture and other forms of failure in mechanical structures [5]. Another notable exception to the lack of emphasis in engineering education has been the discipline of nuclear engineering in which, for example, Northwestern and other universities have offered graduate level courses in reliability engineering to insure that students receiving degrees in nuclear engineering had an in-depth understanding of reliability engineering principles and their application to the design and operation of safe nuclear reactors. In the nuclear engineering case, considerable emphasis was placed on predictive models such an empirical approach was clearly not an acceptable alternative. Even a single major nuclear plant meltdown or release of radioactive materials is unacceptable!

A third phase of reliability engineering has evolved over the last 25 years, with the explosion of the semiconductor industry and the digital computer industry. Semiconductor devices and computer systems have spawned a renaissance in reliability engineering. This has occurred primarily within industrial manufacturing facilities and associated research and development laboratories. Some spillover into universities has occurred with serious graduate programs in reliability engineering at The University of Maryland (focus on electronic packaging, heat transfer, and FEM), The University of Arizona (mechanical engineering), Clemson University (semiconductor reliability), and

a variety of continuing education programs (The University of Southern California, George Washington University, etc.).

During this same period, important contributions to understanding elemental failure mechanisms and precursors have been made in university research. For example, large scale 3D computer simulations are underway at Penn State [9] for use in studies of micromechanical properties, such as the three-dimensional local stresses around individual grains. Figure 1 shows a computer simulated intergranular fracture. It is believed that simulations such as these can be used to successfully study the effects of residual strain, defects, grain size and shape, and other factors on the phenomena such as creep and microcracking which are the precursors to mechanical failure. Progress has been made as well in the development of quantitative non-destructive evaluation methods and to the development of quantitative non-destructive evaluation methods to identify failure have been made by university research. For example, Professor Sachse of Cornell and others have developed acoustic emission methods for fault detection in relatively simple materials and geometries. Excellent reviews of this area are provided by Achenbach and Rajapakse [1] and by Datta, et al. [4]. Current efforts focus on the extension of the concepts to more complex composite materials. The dynamic environment of interest in much condition monitoring remains to be addressed.

University research has also resulted in new and improved sensor concepts and signal processing algorithms which may be useful for detecting and classifying precursors to mechanical system failure. Some of the most promising of these are considered subsequently in this paper. Likewise, a range of new materials processing methods which may contribute to mechanical failure prevention have emerged from university research.

The shortcoming of university research from a reliability and condition monitoring perspective, in the judgment of the present authors, is that in general it has not focused on integrating of these various elements of failure methods, sensors, signal processing, and new materials with the goal of developing improved condition monitoring systems. This integration step would contribute directly to improved condition monitoring systems. Additionally, it would undoubtably uncover new research questions that are particularly relevant to condition monitoring and condition based maintenance.

The University Role for the Remainder of the Decade

For the remainder of this decade, it is imperative that the U.S. university community contribute to mechanical failure prevention in three ways in our judgment. First, the basic and applied research necessary to move from a machinery diagnostic to a machinery prognostic capability is required. Such a capability not only allows one to detect precursors to component or subsystem failure but to predict the remaining safe operational life as a basis for maintenance decisions. Second, new materials and manufacturing methods are now under development which will in all likelihood have different dominant failure modes than currently encountered. These modes must be characterized, predictive capabilities developed, and strategies for cost effectiveness maintenance identified. Third, the fundamentals of reliability engineering must be introduced as a prominent component of the undergraduate engineering curriculum. The

16

first two of these will involve faculty, staff, and graduate students in multidisciplinary research programs, while the third is dominantly an undergraduate rather than graduate education initiative. These three domains of university-based activities in mechanical failure prevention are discussed in detail in the remainder of this paper.

Machinery Prognostics: Research and development of the last five years has resulted in substantial improvements in the ability to detect precursors to failure in mechanical systems such as gear boxes. The challenge remains, however, to detect failure precursors at still earlier times and to fuse this data with model-based information to predict the remaining safe operational life of the component or subsystem. This fusion of model-based information with improved precursor detection methods represents the essence of a prognostics capability. In the judgment of the present authors, it is the next major milestone to be achieved in mechanical diagnostics. The same opinion has been expressed by others working on mechanical diagnostics and failure prevention. At the 1992 International Gas Turbine and Aeroengine Congress in Cologne, .jr example, representatives of Saudi Aaramco, Phillips Norway, KLM Royal Dutch Airlines, and Dow Chemical participated in a panel discussion on the diagnostics for turbomachinery. All of these companies are presently using some form of diagnostics, and in terms of future needs all identified a prognostic capability as a critical goal to be achieved.

The university-based research agenda required to achieve and utilize a prognostics capability has the following components: (a) improved models of failure signatures at the component level; (b) improved sensors; (c) enhanced failure signature detection and classification strategies; (d) fusion of measured data and model based information; and (e) decision methodologies to utilize a prognostic capability to optimize maintainability in terms of cost, safety, or other relevant considerations. Research issues important in each of these contexts are as follows.

(a)   Improved models of failure signature at the component level -- Seminal work on the modeling of gear tooth induced vibrations has been done by W. D. Mark now of the Applied Research Laboratory, The Pennsylvania State University. An excellent overview of this work and its implications is provided in the most recent edition of the *Handbook of Acoustical Measurements and Noise Control* [7]. Most modern spur and helical gears utilize gear tooth geometries that are involute curves or modifications thereof, which ideally transmit an exactly constant angular velocity ratio between meshing gears. In reality, deviations from the ideal tooth contour results in vibratory excitation originating at the meshing teeth of each gear pair. These deviations can arise from such causes as elastic deformation of the gear teeth, machining errors in the contours of individual teeth, tooth spacing errors, and tooth wear. A detailed harmonic analysis methodology has been developed which allows *a priori* prediction of the spectral content associated with these and other deviations in gear configurations from the ideal.

This modeling of gear tooth-induced vibration is directly applicable to prognostics. A measured spectrum can be compared with the predicted one to determine precisely the nature and extent of wear that has occurred and elastic deformation present. Models for fatigue and other modes of failure can then be employed to predict the remaining safe life of the gear system. Additionally, these models for gear tooth vibration can be coupled with mathematical descriptions of incipient failure mechanisms, such as fatigue

17

cracking. When a crack first begins to develop, the effective local modulus of elasticity is reduced, resulting in a slight change in the local tooth geometry under load and a corresponding periodic alteration in the rotational speed or torque transmitted. The development and validation of such combined geometry-failure mechanism models is a critical next step in the development of a prognostic capability, making possible the prediction of the precise character of the failure signature and the remaining safe life of the component. An added benefit of such a model is its potential utility for conducting simulated seeded fault testing at a fraction of the cost associated with the corresponding experiments.

In many applications of condition-based maintenance, the measurements available will be from sensors somewhat removed from the individual component (e.g., rotational speed variations and vibration levels on the outside of a gearbox). Therefore, a second step in model development is the translation of the component level failure signature into that observed at candidate measurement locations, both to determine the required sensitivity and optimum placement of sensors for precursor detection and classification. Existing finite element methods should be adequate to do this component-to-subsystem scale-up in some circumstances, although other methods could be required to accurately predict the small variations introduced by the incipient failure in relation to larger level vibrations and torque variations from other sources and in progressively more complex system types.

During the remainder of the decade we recommend that the university-based mechanics community devote attention to the development of detailed and comprehensive component level models such as already available for gears for other common components such as bearings and shafts. These should be integrated with the best failure mechanism models emerging from the materials science and engineering communities to predict failure signatures at the component level. Additional research is needed as well to translate this component level information into that available at the subsystem or system level. Depending on system complexity, some hybrid of improved finite element methods and system identification methods could in fact be involved. System identification may provide a useful formalism for constructing a transfer function between a perturbation to a given component (caused by a failure precursor) and its manifestation at the system or subsystem level.

(b)   Improved sensors -- Enormous progress has been made in recent years in miniaturizing and ruggedizing a broad range of sensor types. Initial experience has been obtained as well in embedding sensors in mechanical components. An important area of university-based research and development in the coming years is the perfecting of such embedded sensors and evaluation of their utility in mechanical failure prevention. Conceptually the advantage offered by an embedded sensor is being able to place it in close proximity to or actually in the component of interest and thereby avoid the contamination of the signal as it propagates through mechanical paths to more accessible locations (measurement in the gear vs on the outside of the gearbox, for example). In practice, the internal noise field unrelated to a developing fault, reduction in component strength due to the presence of the sensor, or other factors may partially or totally offset the advantages of proximity.

18

Several sensor types are of potential interest. These include solid state strain gauges; piezoelectric and electrostrictive vibration sensors; fiber optic temperature and vibration sensors; and X-ray portable heads with fiber optic cable for real-time remote residual string analyzers to mention only a few. The use of distributed sensors and sensor arrays along with local microprocessors to digest and analyze data in real-time, is also developing at a rapid rate as the computing power and cost of sophisticated chips comes down and such sensor-processor combinations will likely be of importance. The advent of "smart" sensors which not only detect motion or deformation, but then analyze it locally and respond by generating a force back on the object being monitored, while still in its infancy, shows great promise. In this connection, the development of integrated ceramics containing piezoelectric sensor and actuator functions as well as resistive, capacitive, and inductive networks, etc. is notable. Up-to-date reviews of sensors, actuators, and smart versions of them can be found in the work by Cross [2], Newnham [12], and Uchino [15]. We believe that one or more of these new sensor/actuator technologies will be applicable to on board condition based maintenance for the detection of precursor phenomena in mechanical systems, machinery, etc.

Also use of the remote sensing residual stress analyzer of Ruud et al. [14] to monitor the creep rate might be combined with the predictive capabilities of the empirical Voight time to failure equation [16] which uses as input the first and second derivatives of the creep.

(c) Enhanced detection algorithms and strategies -- A successful prognostic capability will require early detection and classification of failure precursors in an inherently noisy environment. Such events could be in the form of one-time transients. Alternatively, they may be manifested in the gradual change in some quantity measured over a long period of time. Finally, the development of some types of failure mechanisms may be more amenable to active than passive detection (i.e., detecting a change in response of the system to an artificially induced perturbation vs monitoring the sensor outputs as they naturally occur). Some of the methodologies that may have a role in these detection and classification problems are wavelet transforms, nonlinear dynamical systems concepts, and the extension of quantitative nondestructive evaluation methods to the dynamic machinery environment.

Wavelet transforms are proving very effective for detecting and classifying transient events. The wavelet transform differs from other short-time transforms, such as the short-time Fourier transform, in that it has a constant time-bandwidth product; or in other words it is a constant Q filter. Much of the current interest in wavelets may be credited to Grossmann and Morlet, who developed the first practical method for computing the wavelet transform. The method is shown schematically in Figure 2. The Fourier transform is first computed and then multiplied by a scaled window function and a translation operator in the transform domain. Since the Fourier transform has both modulus and phase, the same is true of the wavelet transform computed with this method. Also note that the bandwidth of the window function increases with frequency to achieve the constant Q characteristic. The wavelet transform of white noise computed in this way and that of a signal with two superimposed transients is shown in Figure 3. Here the amplitude of the transform is indicated by a gray scale. Of particular note is the sensitivity of the phase of the wavelet transform to a transient so small that it can barely be seen on the time trace. It is just this sensitivity to transients that makes the

19

wavelet transform of practical utility in detecting one-time changes in system state that are precursors to mechanical system failure.

Several subsequent developments in the implementation of wavelet transforms may be of particular importance in their applications to mechanical diagnostics. For example, Zhong [19] has developed an efficient method of computing the discrete wavelet transform directly, rather than through initial computation of the Fourier transform. More importantly, they have shown that the components of the discrete wavelet transform at each temporal scale can be replaced by delta functions at the transform maxima and minima and still retain the essential information about the function. This discovery has potentially important data compression implications for diagnostics applications, perhaps allowing much longer temporal records to be both stored and analyzed for transient events than would otherwise be possible. A second recent development is the cross wavelet transform by Young [18]. In the same way that the cross spectrum often yields important information about the relationship of two periodic signals not available from a direct comparison of two power spectra, so also does the cross wavelet transform have the potential for providing information not otherwise available about transient events through analysis of simultaneous outputs from two sensors. A third is the use of wavelet based higher order spectra developed by Wilson and co-workers [17]. Work to exploit all of those advances in wavelet based detection and classification for improved mechanical diagnostics is needed in the academic community, extending the work of Li [11].

It is reasonable to expect that improved mechanical diagnostics will required both improved detection and classification of events with transient and with continuous signatures. Whereas wavelet based methodologies are promising for transients, nonlinear dynamical systems concepts may be advantageous in the context of continuous events with broadband signatures. Nonlinear dynamical systems, also called chaotic systems in the literature of the past few years, can be generated by relatively simple sets of equations. More important from a diagnostics perspective is the fact that while they have broadband power spectra, each has a unique pattern or geometry when viewed in phase space or with a simple time delay mapping $(X(t + T)$ as a function of $X(t))$. A simple example is the van der Pol equations, for which the power spectrum and a time delay map for one dependent variable are shown in Figure 4. From the point of view of mechanical diagnostics, nonlinear dynamical systems concepts are potentially important because phase space characterization may provide more information on latent failure development at an earlier stage than available in the spectral domain. In practice, one would not in general visualize the phase space plot of the output of a sensor, but rather monitor the trend in some parameter which characterizes the plot. Candidates include the largest Lyapunov exponent and the linear intrinsic dimension, both of which are indicative of the number of independent variables required to characterize the system.

Within the mechanical diagnostics context there are at least two major research issues that must be addressed to realize the full potential of nonlinear dynamical systems concepts. First, improved methods are needed for characterizing the nonlinear system or changes therein in a noisy environment. At the present time, the phase space plot can be successfully reconstructed from a noisy signal (signal-to-noise ratio or order unity) only with significant *a priori* information about the nonlinear system. Second, methods

20

need to be developed for utilizing the power of nonlinear system concepts for prognostics as well as diagnostics. Significant university-based activity will be required to answer both of these challenges.

More generally, mechanical diagnostics depend largely on a "passive" detection and classification strategy, in that sensor outputs are monitored and analyzed as they naturally occur. An alternative "active" strategy is to periodically perturb the system in a known or measured way and analyze the system response to those perturbations. Such active techniques have been quite useful in quantitative nondestructive evaluation in a static or quasi-static environment but have yet to be thoroughly investigated in the dynamic machinery condition monitoring context. The potential promise of such techniques in this new application is nevertheless suggested by the extreme sensitivity to system state of the system transfer function in initial experiments conducted at the Applied Research Laboratory of Penn State. Continued research in the academic environment is needed to define the limits of utility of such active approaches and to develop optimal methods for combining active and passive approaches.

(d)   Model-data fusion -- Significant developments have occurred in recent years in the combination (fusing) of data from different sources to provide a more comprehensive representation than provided by any one of the sources. An excellent overview of these developments is given in the Proceedings of the 1991 Joint Service Data Fusion Symposium [8]. The application of some of these methods to mechanical diagnostics may be beneficial for the utilization of data from several different sensor types. Data fusion research has not in general focused on the optimal combination of sensor based and model based information; however, and it is just this fusion problem that is critical in the development of a prognostics capability. Strategies for optimal model-data fusion appropriate to the prognostics problem are an important undertaking for the academic community. At the outset, this research might focus on fatigue cracking in gears, for which both models and data are either in hand or under development. The essential question to be addressed by this research is how models and data may be combined to overcome the inherent limitations of each to provide an acceptable prognostic capability. The limitations on data quality and model fidelity will vary depending on the application, but neither will in general be as good as one would desire as the basis for a prognostic capability, and this shortfall must be made up by the synergistic manner in which the two types of information are brought together.

(e)   Decision methodologies -- An optimal maintenance plan for a given application requires integration of prognostics information, user specified cost function, and constraints on maintenance actions. The cost function might involve maintenance cost, down time cost, safety, or some combination of these and other factors. Practical limitations on the maintenance plan could take the form of a system being available for maintenance only during certain time intervals dictated by operational constraints. The academic research challenge will be to integrate prognostics information with constrained optimization methods in such a way as to fully account for all of these factors. It is our current thinking that new decision or optimization methods will not need to be developed for this purpose. Rather, extensions and combinations of those already developed to the maintenance context will be involved.

21

Influence of New Materials and Processing Methods: Recent trends in the design of "engineered materials" having properties which can be tailored by close control of their microstructures, both in the bulk and on the surface, are expected to impact mechanical reliability. Parts made from such engineered materials can have unique properties not previously possible. Also, new methods of processing more traditional structural materials can lead to greatly altered mechanical properties. An example of the former is the ausforming process whereby precision gears and other metal parts can be made to net-shape without the necessity of grinding. The fatigue and failure of the "ausformed" gears, bearings, etc. is going to be markedly different from the traditional parts which were machined or ground to shape. Chemical machining of brittle ceramics and composites while still largely an exploratory method could also have a major impact on the reliability of "engineered" materials.

Recent advances in sophisticated multi-layer coatings of ultra-hard, high-temperature coatings of diamond, Ti, TiC, etc., by production techniques such as physical e-beam deposition also promise to improve the mechanical reliability of rotating machinery. Production facilities capable of depositing 15 hgm/hour of such coatings on turbine blades, gears, etc. are now in existence. Most uses so far have been for military and aerospace applications, but should now reach out into the commercial sector.

While such advances in materials and processing methods take many years to implement, because of the large capital investments and the learning process involved, we feel confident that they will result in major improvements in the reliability of parts and machinery.

Reliability Engineering in the Engineering Curriculum: The ABET approved course requirements for engineering still do not include courses for the sound training of engineers in the basic principles of reliability engineering! This is, in our opinion, a serious deficiency in the present engineering curricula. Engineering literature is filled with articles containing basic errors in the applications of reliability engineering to the solution of specific problems. Since the fundamental relationships of reliability engineering are generally applicable to all branches of engineering -- mechanical, electrical, civil, chemical, environmental, etc. -- it is our thesis that they should be taught at the undergraduate level to all engineering students as a required course. This is essential in condition-based maintenance where one is dealing with a complex variety of reliability issues including monitoring precursors, signal analysis, sensor and electronic circuit board reliability, maintenance logistics and economics, and reliability modeling to obtain accurate predictions of the time to failure. The latter will probably require a detailed grasp of the physics of failure since this is what governs the precursors and the statistics of the failure occurrences. This is a multidisciplinary undertaking even within a College of Engineering. Table 1 shows a chart with some of the disciplines involved in reliability engineering.

Table I. Interdisciplinary Nature of Reliability Engineering

ENGINEERING					
Mechanical	Chemical	Electrical	Industrial	Civil	Computer
finite element analysis vibrational analysis micromechanics fatigue, creep fracture analysis tribology, wear	process control corrosion tribology lubrication	semiconductor failure analysis power system reliability rotating machinery signal processing sensors accelerated life testing	reliability modeling SPC NDT residual stress analysis human factor reliability	construction studies concrete structures asphalt bridge design	software reliability system reliability CAD fault tolerant design databases

SCIENCE			
Physics	Chemistry	Math	Materials Science
solid state physics physics of failure mechanisms chaos theory	electrochemistry stress corrosion cracking corrosion	applied statistics statistical topology homogenization theory	microstructural analysis/ micromechanics metallurgy structure/property modeling materials synthesis and processing computer simulation characterization tools, TEM, etc.

Conclusions

The U.S. University community can take two steps to enhance its contribution to mechanical failure prevention, in our judgment. First, curriculum changes are needed to provide undergraduate engineering students with a solid foundation in reliability engineering. A multidisciplinary course sequence presenting the practical elements of mechanical failure detection and control should be designed and offered as a joint undertaking by mechanical, electrical, chemical, and industrial engineering departments working together with materials science and engineering departments. Second, there is a great need for the establishment of interdisciplinary graduate research programs in reliability engineering. Such programs would provide core support for a concentrated research effort aimed at developing a prognostics capability for systems of the future. Such a capability has a broad range of economic and safety implications in both the civilian and defense sectors of the United States.

References

[1]   Achenbach, J. D. and Y. Rajapakse (editors). Solid Mechanics Research for QNDE. Martinus Nijhoff, Boston (1987).

[2]   Cross, L. E. Proc. International Summer School on Ferroelectrics, p. 1-85 (1991). Ascona, Switzerland.

[3]   Cruse, T. A. (editor). Reliability Technology, AD-Vol. 28. American Society of Mechanical Engineers, New York (1992).

[4]   Datta, S. K., et al. (editors). Elastic Waves and Ultrasonic Non-Destructive Evaluation. North Holland, New York (1990).

[5]   Freudenthal, A. M. (editor). International Conference of Structural Safety and Reliability. Pergamon Press, New York (1972).

[6]   Grossmann, A. and J. Morlet. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape. _SIAM Journal of Mathematical Analysis_ 15:723-736 (1984).

[7]   Harris, C. M. (editor). _Handbook of Acoustical Measurements and Noise Control_. Third Edition. McGraw-Hill Inc., New York (1991).

[8]   Kessler, O. _Proc. Joint Service Data Fusion Symposium_ (1991). Warminster, PA, Naval Air Development Center.

[9]   Kumar, S., et al. _Journal of Statistical Physics_ 67:523-551 (1992).

[10] Lewis, E. E. _Introduction to Reliability Engineering_. John Wiley, New York (1987).

[11] Li, C. J. and Jun Ma. Bearing localized defect detection through wavelet decomposition of vibrations. In _Economic Implications of Mechanical Failure Prevention_, pp. 53-62. Edited by H. C. Pusey and S. C. Pussey. Vibration Institute, Willowbrook (1992).

[12] Newnham, R. E. and G. R. Ruschew. _J. American Ceramic Society_ 74(3):463-480 (1991).

[13] Pollino, E. _Microelectric Reliability_, Vol. 1. Integrity Assessment and Assurance House, New York (1989).

[14] Ruud, P. S., et al. _Advances in X-Ray Analysis_ 27:273-283 (1992).

[15] Uchino, K. _Proc. First European Conference on Smart Structures and Materials_, pp. 177-180 (1992). SPIE

[16] Voight, B. _Science_ 243:200-203 (1989).

[17] Wilson, G., Private Communication.

[18] Young, R. K. _Wavelet Theory and Its Applications_. First Edition. 223 pp. Edited by Jonathan Allen. Kluwer Academic Publishers, Boston, MA (1993).

[19] Zhong, S. Complete Signal Representation with Multiscale Edges. New York Technical Report No. 483 (1989).
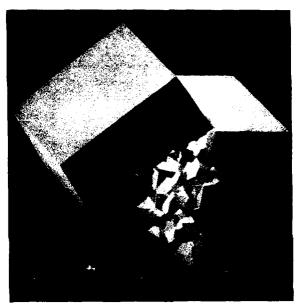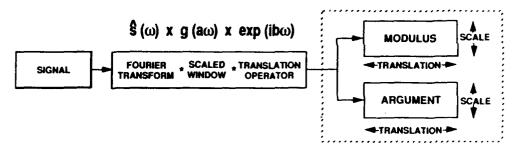
Figure 1.   Computer simulation of intergranular fracture.

$$\hat{s}(\omega) \times g(a\omega) \times exp(ib\omega)$$

SCALED WINDOW (SHIFTED GAUSSIAN)
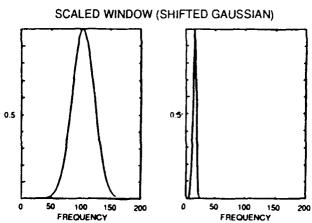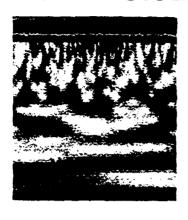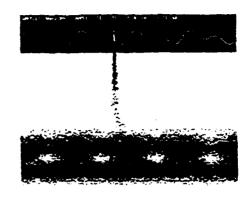
Figure 2.   Schematic representation of the continuous wavelet transform calculation.

WHITE NOISE   TRANSIENTS
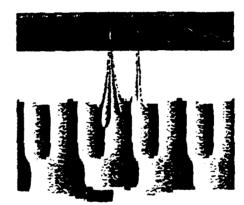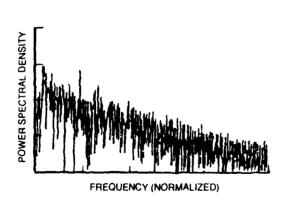
Figure 3.   Wavelet transforms of white noise and transient signals.
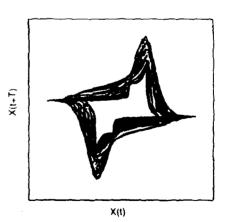
Figure 4.   Power spectrum and time delay representation of the van der Pol equation.

26

SHAFT CRACK DETECTION IN ROTATING MACHINERY

Donald E. Bently
Bently Nevada Corporation
Minden, NV 89423

ABSTRACT: The three major mechanisms correlated with a shaft crack in rotating machinery and basic rules for vibration monitoring to detect cracked shafts at early stages are outlined. The effects of start up and shut down on the 1x and 2x vibration are emphasized. The role of misalignment in shaft cracking and balancing if a cracked shaft is suspected are discussed. Horizontal and vertical machines are mentioned. Recommendations are provided.

Key Words: Synchronous (1x); twice per revolution (2x); vibration monitoring; slow-roll; trend analysis; transient; amplitudes; split of natural frequencies; misalignment; balancing; radial preload. snap action; top dead center.

Introduction: Life of machinery can be significantly extended by undertaking corrective actions as soon as any machine malfunction occurs. Critical machinery should be monitored continuously, and vibration analysis techniques should be applied.

The three major mechanisms correlated with a shaft crack are:

1. When a shaft develops a crack, it nearly always bows. This creates additional unbalance. The additional unbalance generates another synchronously rotating centrifugal force which excites the rotor synchronous (1x) response. Since there is always residual unbalance in the rotor, new unbalance will be added vectorially; therefore, the 1x response will be different than before the crack occurrence and changes as the crack propagates.

2. A cracked shaft's lateral stiffness becomes asymmetric, that is, the shaft is weaker in the direction of the crack. With the existence of a shaft radial preload (due to any force perpendicular to the shaft axis, such as gravity force in horizontal shafts, misalignment forces, and/or fluid flow-related forces), the shaft will respond with vibrations with a frequency twice-per-revolution (2x). Very often shafts are originally somewhat asymmetric by design (such as two-pole generators) or by assembly imperfections (such as non uniformly clamped disks or keyway slots). The shaft transverse crack will modify the original rotor asymmetry. Thus, it will modify the shaft vibrational 2x response component with or without a crack.

3. When three separate phenomena occur simultaneously: (1) a shaft symmetry (a crack or a keyway, or a notch or whatever); (2) a radial sideload (such as misalignment or pumping sideload);(3) a lateral resonance at or near twice operating speed (2x);the rotor system will generate and propagate a crack very rapidly. These conditions must not be allowed simultaneously on any machine.

<u>Vibration Monitoring:</u> These three physical mechanisms which are correlated with a shaft transverse crack imply using vibration monitoring systems for crack detection. The basic rules for vibration monitoring in order to detect shaft cracks at early stages are:

1. Observe the synchronous (1x) component of the rotor response. Any change in either the 1x response vector amplitude or phase which is not explained by load or other normal operational parameter changes may indicate shaft crack development.

2. Observe slow-roll, 1x and 2x vector changes and also the shaft static position (shaft centerline) changes. Any of these may indicate the existence of a shaft crack. Shaft static position (dc gap) should be observed not only at rest, but what is more important, at the operating speed.

3. Observe the twice-per-turn (2x) component of the rotor response. Any change in the 2x response vector in either amplitude or phase may indicate shaft crack development. These changes may occur smoothly, or by sudden jumps.

The method of early detection of shaft cracks is based on the trend analysis -- both steady-state, "on line," direct observation of 1x and 2x vibrational response vectors (Fig. 1) and indirect comparison of occasionally recorded "transient" responses. The limits such as "alert" for warning, and "danger" for alarm should be specified for each machine according to its operational conditions.

Amplitude and Phase versus Time (APHT) is an acronym used to describe the *trend plot* of amplitude and phase versus time. The APHT was developed to assist in interpreting the amplitude and phase data. This data may be presented in both *Cartesian* and *polar* formats. Commonly used for 1x and 2x vibration data.(Fig. 6)

If only vibration amplitudes are taken into account, a 25 percent increase or decrease about the normal level should be considered as a warning, 50 percent increase or decrease as a major alarm. Note, however, that the real problem may lead to more dramatic changes in phase than in the amplitude; that is why both amplitudes and phase trend should be observed. Similar percentage values for phase acceptance, warning and alarm limits should, therefore, be specified.

<u>Recommendations for Early Detection of Shaft Crack Using Vibration Monitoring Data:</u> There are a few enhancements to the above specified three rules that are associated with specific dynamic conditions of machine operation; they are as follows:

● Rotative speed: Effect on 1x vibration.

It is well known that the rotor synchronous (1x) response depends not only on the amount of unbalance but even more on the frequency at which this unbalance rotates, i.e., on the rotor rotative speed. When the speed is in the vicinity of the natural frequency of any bending mode of the rotor, the response may increase dramatically. This effect is called resonance. The machine operational speed is presumably chosen in a nonresonant range of speeds, below the first, or in between two adjacent widely spaced natural frequencies. In such range of rotative speeds any change in the unbalance amount will produce only small changes in the 1x response vector. This amount may be so small that it can go unnoticed, hidden in the instrumentation noise level.

A different situation takes place when the rotor passes through the resonant ranges of speeds during start-up or shutdown of a machine operating above the first balance resonance. During these transient conditions, the response at resonant speeds is relatively high; therefore, the 1x response vector modifications due to shaft cracks are more easily detectable (Figs. 2A--2B).

28

The histories of many saves of cracked shafts of several manufacturers, with data from protection monitoring is that more than 75% of all instances have rotative speed (1x) amplitude or phase changes only. The remaining (less than 25%) are predominantly nuclear reactor coolant pumps, many of which have a lateral (pump load).resonance at or near 2x.

Recommendation: Record and document both the amplitude and phase of the 1x vibrational data during each start-up and shutdown, and compare them with the previously recorded ones.

- Rotative speed: Effect on 2x vibration.

Similar rules of response magnification in the resonant range of speeds apply in the case of rotor 2x vibrational responses, except that the corresponding resonant rotative speed ranges occur at about half value of these for 1x vibrations.

Recommendation: Record and document both the amplitude and phase of the 2x vibrational data during each start-up and shutdown, and compare them with the previously recorded ones.

- Transient processes: Start-up and shutdown.

During transient processes of start-up and shutdown of the machine, the torque conditions are different. This may affect the rotor lateral vibrational responses, especially when there is a strong coupling of modes. The "normal" differences in these responses are usually known for each particular machine design. A shaft crack may dramatically change the situation, and increase the start-up and shutdown response differences since it affect the integrity of the machine, which can be highly torque-sensitive.

Recommendation: Compare the corresponding start-up and shutdown data in order to detect increasing differences in rotor vibrational responses.

- Shaft crack-related split of natural frequencies.

A shaft transverse crack causes an asymmetry in the rotating system. This results in differences in the horizontal and vertical vibrational responses of the rotor, especially noticeable at resonant speeds. One peak for symmetrical system response (no crack) will split into two adjacent peaks. The effect, however, is highly sensitive to the amount of damping in the system; thus it may be unnoticeable.

Recommendation: Watch for the appearance of "split resonances" in the start-up/shutdown 1x data. They may indicate a cracked shaft.

- Decrease of values of the rotor natural frequencies.

A shaft crack affects the rigidity of the rotor system causing a decrease of the natural frequencies of the shaft bending modes. The amount of the decrease is not the same for all natural frequencies, but depends on a particular mode-versus-crack location. Some modes and the corresponding natural frequencies may be unaffected; some of them may exhibit noticeable changes. It requires a very large crack to make a noticeable decrease in resonant frequencies.

Recommendation: Watch for decreases in the rotor natural frequencies of bending modes. If the resonant peaks of 1x or 2x (and even 5x or 10x on impellers with 5 vanes) responses occur at gradually lower speeds, this may indicate that the shaft is severely cracked. We pioneered the observation of 2x polar plots and 5x (vane passage) polar plots for the observation of the lateral resonances higher than operating speed. These are vital for cracked shaft studies.

- Choice of operational speed.

It is well known that the rotor operational speed should not be chosen in the vicinity of any rotor system natural frequency. This prevents resonant amplification of the 1x unbalance-related vibrations at operating speed.

It is absolutely important that the rotor not operate in the vicinity of the resonance of 2x vibrational component, thus the rotative speed should not be chosen at or near a half of any balance resonance speed in order to prevent resonant amplification of the 2x component. Any machine with this characteristic that is in critical or vital service should be modified to avoid this on a high priority basis.

The rule can eventually be completed by the recommendation not to operate the machine at 1/3 of any balance resonance in order to avoid 3x component resonance. This applies, however, only to machines with very poor effective quadrature dynamic stiffness (damping).

Recommendation: In the choice of operational speed, it is important to remember that a shaft crack causes a decrease in the rotor natural frequencies due to reduced system stiffness. If the operating speed is chosen only slightly lower than any 1x or 2x resonant speed, then the propagating shaft crack can cause the machine to operate at a resonance, a most unfavorable condition, which would accelerate the shaft crack propagation even more.

- Role of misalignment and Radial Pumping Sideload in shaft crack development.

Misalignment and radial pumping sideload of rotors in machine trains are considered to be major contributors to shaft cracking. Nowadays, due to widespread vibration monitoring use, the standing philosophy considers machine vibrations to indicate unacceptable machine performance. Vibrations themselves do not directly cause shaft cracks. The cracks occur due to shaft stress and deformation. When the rotors are operating misaligned, vibrations may decrease; however, stress in rotors increases significantly, leading to cracks. This fact may go unnoticed if only "vibrations" per se are considered.

Recommendation: Record and document the rotor centerline positions during operation (dc gap). The information serves to verify proper alignment condition and to prevent shaft cracks.

- Balancing when rotor is cracked.

An increase of the rotor 1x (synchronous) response is usually correlated to an increase of unbalance. Thus, the conventional cure of the problem calls for balancing. Extreme caution has to be applied in the balancing procedure if a shaft crack is suspected. If the rotor response to calibration weights is not normal and/or erratic, there is a high probability that the shaft is cracked. Repeated attempts to balance a cracked rotor may adversely affect an already serious situation.

Recommendation: Record and document the balance weights, their location, rotor 1x response and slow-roll vectors for each balancing run, and compare against previous data. Do not continue the balancing procedure if the data indicate an unusual behavior of the rotor.

- Horizontal versus vertical machines.

Some 2x lateral vibration component in the rotor response may be commonly present in the rotor response. Most often it is generated by the 1x response force restrained by the nonlinear characteristics of bearings and seals when a steady sideload also exists. (Fig. 3). It often has lower amplitude than 1x component. The 2x vibrations (Fig. 4) in this case are generated from 1x vibrations (Fig. 5).

However, the 2x response due to a shaft asymmetry is generated independently of the 1x response as a function of the stiffer (uncracked) portion of the shaft passing "top dead center" of a steady sideload. As

30

this occurs a "snapping" action of the rotor would follow, which often rapidly propagates an already existing crack.

Two factors are necessary to induce an independent 2x vibration component in rotating machines: asymmetry in the rotating system (for instance, due to shaft crack) and a radial preload applied to the shaft (a force perpendicular to the shaft axis of rotation). In heavy horizontal machines this radial preload is somewhat "natural" and corresponds to the rotor weight. That is why historically the 2x vibration component is associated with gravity ("gravity resonance at half speed"). The gravity force in horizontal machines is not, however, the only possible radial force. There exist other forces applied to the shaft in radial directions and not necessarily collinear with a gravity force. These forces are originated by rotor misalignment in machine trains and by flow in fluid handling machines. Both types of forces can be significantly high and their magnitudes may exceed the magnitude of the gravity force. The implications of this fact are obvious: (i) Radial preload can exist in horizontal, as well as vertical machines. (ii) In horizontal machines the resultive radial preload is not necessarily downward vertical. The recommended methods for early detection of shaft cracks apply, therefore, to both types of rotating machines, horizontal and vertical.

Final Remarks: The important points in the plan to prevent catastrophic failures from cracked rotors in rotating machinery are: the ability to understand the factors contributing to crack propagation, and to detect the shaft crack existence during machine operation. Case histories and experimental studies indicate that with proper vibration monitoring and signal processing, the catastrophic damage caused by cracked rotors could be drastically reduced.
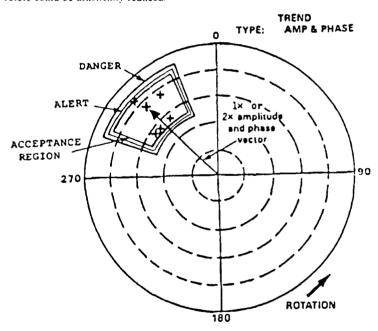
FIG. 1 ACCEPTANCE REGION FOR 1X AND 2X RESPONSE VECTORS IN THE POLAR PLOT FORMAT ACCEPTANCE REGIONS CAN BE DEFINED WITH VARIOUS BOUNDARIES, DEPENDING UPON THE VIBRATION CHARACTERISTICS OF EACH MACHINE UNDER ALL NORMAL OPERATING CONDITIONS

31
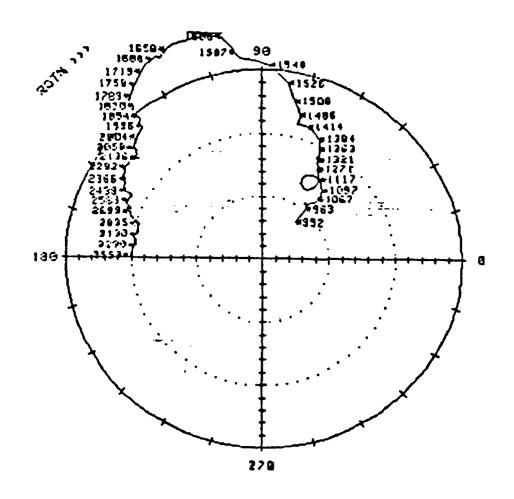
FIG 2A POLAR PLOT OF SYNCHRONOUS RESPONSE (1X) OF A CRACKED
TURBOGENERATOR ROTOR. (SEE FIG. 2B FOR 2 DAYS LATER).
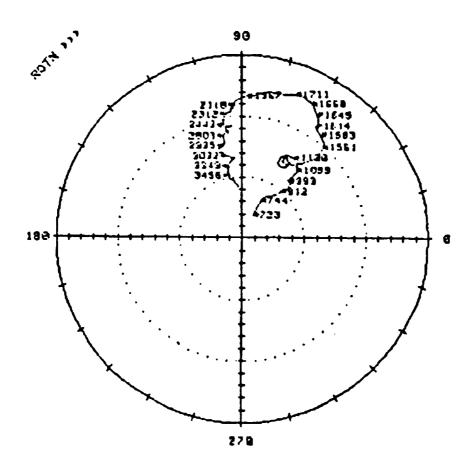
FIG 2B   POLAR PLOT OF SYNCHRONOUS RESPONSE (1X) OF A CRACKED
         TURBOGENERATOR ROTOR. (SEE FIG. 2A FOR BEHAVIOR 2 DAYS EARLIER)
         NOTE A DECREASE OF 1X RESONANT AMPLITUDE AND A PHASE SHIFT.

33
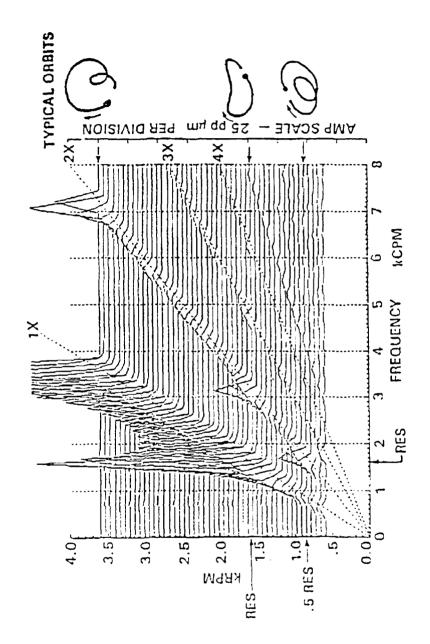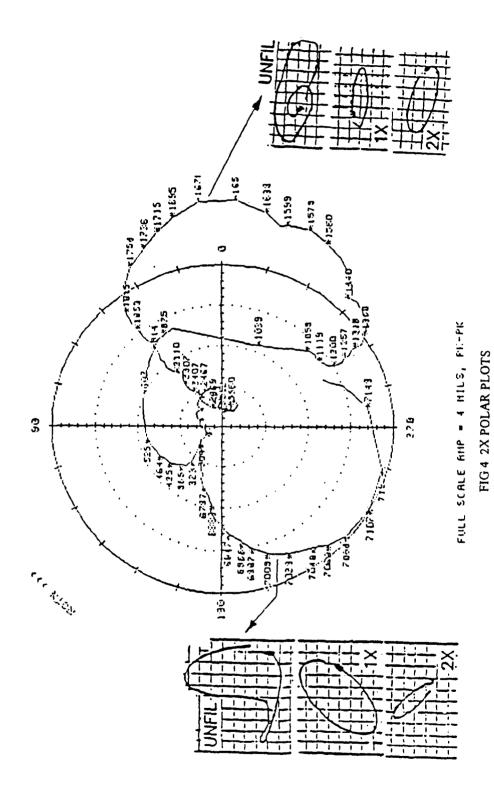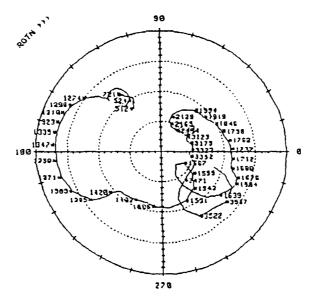
FIG 3 SPECTRUM CASCADE (WATERFALL) PLOT OF VIBRATION DATA
FROM A TURBOGENERATOR SET.

34

FULL SCALE AMP = 4 MILS, FL-PK

FIG 4 2X POLAR PLOTS

FULL SCALE AMP = 10 MILS pp

FIG 5 1X POLAR PLOTS
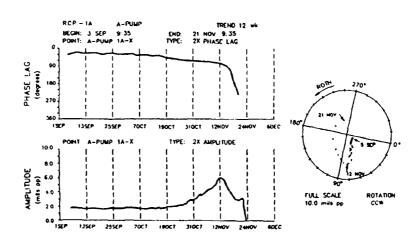
Fig 6 The Amplitude Phase vs Time plot. This is a trend plot of nX Amplitude and Phase, with x=2 in this figure.

ARE MODES A USEFUL DIAGNOSTIC
IN STRUCTURAL FAULT DETECTION?

Mark H. Richardson
Vibrant Technology, Inc.
Jamestown, CA

Abstract: Modal testing has become commonplace in many industries as an R&D tool, and for trouble shooting noise and vibration problems in operating machinery and equipment. Very little use has been made of this technology, however, for detecting structural faults or defects in machines and structures. A structural fault, such as cracking, delamination, unbonding, loosening or wear out of fastened parts, etc., will cause changes in the measured vibration response of a structure. These changes will, in turn, cause changes in the structure's experimentally derived modal parameters.

Using existing modal testing technology, a structural monitoring system which measures the vibration of a structure, identifies changes in its modal parameters, and predicts occurrences of structural faults could be built. Such a system would provide a level of accuracy far beyond the traditional peak picking implementations of the past. Also, its implementation can benefit from a complete a priori knowledge of the structure's dynamic characteristics, which is contained in its modal properties.

In this paper, several important issues associated with the use of experimentally derived modal parameters as a means of detecting structural faults are examined. Also included are some experimental results which demonstrate how modal parameters are changed by simulated faults.

Introduction: The underlying principle behind this fault detection method is that vibration is a *sensitive indicator* of changes in the physical integrity of any mechanical structure. When a structural fault such as cracking, delamination, unbonding, and loosening of parts occurs, this causes a decrease in stiffness, (and perhaps an increase in damping), in a local region of the structure. This change in the local stiffness and damping properties directly affects the manner in which the structure will vibrate when excited by applied forces, (either ambient or artificially applied). A common example of this is the bell. If a bell is cracked, then when it is struck, it will give off a more heavily damped "thud" sound rather than the expected lightly damped ringing sound.

The mass, stiffness, and damping properties of a structure determine how it vibrates. Vibration is caused by an exchange of energy between the mass (or inertial) properties and the stiffness (or restoring) properties of a structure. Damping in a structure dissipates vibrational energy, usually as friction heat.

The equations that describe the vibration of a structure are commonly derived by applying Newton's second law to all of the degrees of freedom (DOFs) of interest on the structure.

37

For an experimental situation, this results in a finite set of equations, one for each measured DOF:

$$[M] \{x''(t)\} + [C] \{x'(t)\} + [K] \{x(t)\} = \{f(t)\} \tag{1}$$

where: t = time variable (seconds).

n = number of measured DOFs.

[M] = (n by n) mass matrix (force/unit of acceleration).

$\{x''(t)\}$ = acceleration response n-vector.

[C] = (n by n) damping matrix (force/unit of velocity).

$\{x'(t)\}$ = velocity response n-vector.

[K] = (n by n) stiffness matrix (force/unit of displacement).

$\{x(t)\}$ = displacement response n-vector.

$\{f(t)\}$ = excitation force n-vector.

Notice that the excitation forces and responses are functions of time (t), and that the coefficient matrices [M], [C], and [K] are constants. This dynamic model describes the vibration response of a linear, time invariant structure, subject to any number and kind of externally applied forces, represented by the force vector $\{f(t)\}$. Notice that all solutions to equation (1) are directly influenced by the mass, stiffness, and damping properties of the structure. If the structure is struck with an impulse, such as in the case of striking a bell, equation (1) will yield the impulse response of the structure as a solution. The impulse response of a bell is, of course, its ringing sound. The boundary conditions (mountings) of a structure also influence its vibrational response. This certainly agrees with our intuition and experience. A cantilever beam will vibrate differently than a beam without a fixed end.

Equivalent Representations of Structural Dynamics: In addition to its differential equations of motion given in equation (1), the linear dynamics of a structure can also be represented in several other equivalent forms, as shown in Figure 1. Frequency Response Functions (FRFs), Impulse Response Functions (IRFs), and the structure's modal parameters each fully represent the dynamics of a structure also. Consequently, Figure 1 indicates that if any of the mass, stiffness, or damping properties of the structure should change, we can expect that its FRFs, IRFs, and modal parameters will change also. Conversely, if the measured FRFs, IRFs, or experimental modal parameters of a structure were to change, we can expect that some of the mass, stiffness, or damping properties should have changed also.

In summary then, the modal properties of a structure are directly related to its mass, stiffness, and damping properties. Therefore, changes in the structure's mass, stiffness, or damping properties will cause changes in its modal properties (modal frequencies, modal damping and mode shapes). Changes in the structure's boundary conditions (mountings) will also change its modal parameters.

38

Differential Equations

$M \ddot{x}(t) + C \dot{x}(t) + K x(t) = f(t)$

Fourier Transform

System Matrix

$[M\omega^2 + C\omega + K] X(\omega) = F(\omega)$

Eigensolution → **Modal Parameters**

Matrix Inversion

FRF Matrix

$X(\omega) = [H(\omega)] F(\omega)$

Curve Fitting → **Modal Parameters**

Fourier Transform

Impulse Responses

$[I(t)]$

Curve Fitting → **Modal Parameters**
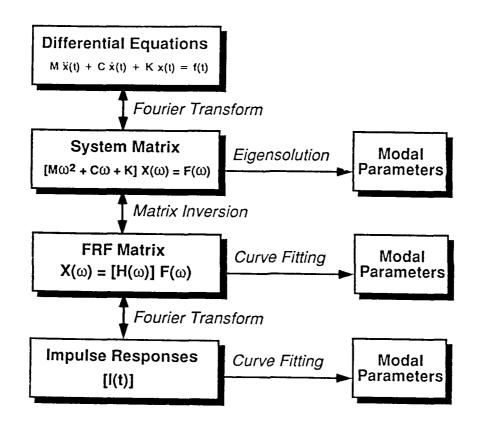
Figure 1. Equivalent Representations of Structural Dynamics.

- Visual Inspection
- Magnetic Field
- Eddy Current
- X-Ray
- Ultrasound

- Acoustic Emission
- Thermal Contours
- Laser Interferometry
- Strain Gage

Figure 2. Traditional NDT Methods.

Anyone who has performed a modal test has probably experienced the strong sensitivity of modal parameters to physical changes in the test setup. Mass loading, ambient temperature changes, and vibration induced changes in the constraints or material properties of the test structure will often cause changes in its measured modal parameters, thus invalidating the test results. The idea behind this approach to structural fault detection , then, is to exploit this strong coupling between changes in a structure's physical properties and its modal properties.

Advantages of Modal Testing as an NDT Method: A wide variety of different Non-Destructive Testing (NDT) Methods have been implemented on structures, as shown in Figure 2. Vibration testing is conspicuous by its absence. Why is vibration measurement (in particular modal testing) advantageous as an NDT method?

1. Modal Testing Can Be Applied to Composite Structures
Modes can be measured on any structure that can be made to resonate. Any structure that is so constructed that acoustic energy can readily travel within its boundaries, will resonate. Therefore, it can be vibration tested. This includes complex structures that contain dissimilar materials such as plastics, graphite epoxies, ferrous and non-ferrous metals.

2. Modes are Sensitive Indicators of Physical Changes.
It is well known among experimentalists who are familiar with modal testing that modes are very sensitive indicators of any changes in the physical (mass, stiffness, or damping) characteristics, or physical constraints (boundary conditions) of the test structure. Ambient temperature changes of a few degrees can often change the stiffness of the structure, which will cause a measurable shift in its modal frequencies.

3. Modes Can Localize a Fault.
Changes in the higher frequency modes of a structure can be used to localize structural faults. *Modes shapes* are acoustic standing deformation waves within a structure. The mode shapes of the lower frequency modes (called fundamental modes) normally cover the entire span of the structure's surface. On the other hand, the mode shapes of the higher frequency modes typically become more localized to particular regions of the structure. (For this reason, they are called local modes.) Therefore, a detected change in a local mode can be used to localize a structural fault.

4. Faults Can Be Detected in Unmeasured Regions of the Structure.
Due to the *global nature* of modes, measurements do not have to be made directly at a fault location in order to detect the fault. *Most other NDT methods require that a measurement be made at the fault location in order to detect it.* Modal frequency and damping of the lower frequency fundamental modes can be estimated from measurements made anywhere on the entire surface of the structure. Frequency and damping estimates of the higher frequency local modes can be obtained from measurements made anywhere in the local region where they are defined.

5. Remote Measurements Using Non-Contacting Transducers Can Be Made.
Vibration is manifested on the surface of a structure. Mode shapes, standing acoustic waves which deform the structure, can be measured with any transducer that measures

40

surface motion, typically normal to the surface. Non-contacting transducers such as photonic sensors and laser transducers can be used to measure vibration.

6. Only a Small Number of Measurements are Required

Only a small number of measurement points (ideally only one) are required to monitor the modal parameters of a structure. Modal properties are typically estimated from FRF measurements. (In certain circumstances, it may be more advantageous to estimate them from IRF measurements.) An FRF is a 2-channel measurement, involving two simultaneously sampled signals; a response signal and an excitation (force) signal.

<u>Controlled Excitation Versus Operating Data:</u> Modal properties are independent of structural excitation. A key difference between operating deflection shapes (forced vibration under different operating conditions) and mode shapes is that operating deflection shapes change with the structural excitation; mode shapes do not. The excitation force(s) are usually not measured when operating data is acquired. (*See* reference [7] for a explanation of the relationship between operating deflection shapes and mode shapes.) To identify modal properties, it is preferable to artificially excite the structure, measure the excitation force(s) and response(s) to form FRFs, and not use operating data.

7. A Wide Variety of Excitation and Signal Processing Methods Can Be Used

Advances in FFT-based test equipment and frequency domain parameter estimation (curve fitting) methods have significantly improved the accuracy and repeatability with which modal parameters can be identified from test data. Modern modal testing methods include the use of;

- Multiple exciters and a wide variety of excitation signals, including many variations of transient, sine, and random signals.
- Multi channel data acquisition and MIMO (multi input multi output) digital signal processing using the FFT (Fast Fourier Transform).
- Multiple reference (Poly Reference) curve fitting of the measurement data to estimate the modal parameters.

8. Modal Testing is Non-Destructive.

Modal parameters can be estimated from FRF measurements that are made using very low levels of excitation, thus incurring little risk of inadvertently damaging the structure during testing. *Sine wave excitation at the structure's resonant frequencies, which can potentially damage the structure, is not required.* Any one of a variety of popular *broad band* excitation signals can be used instead. There are other FFT-based signal processing benefits to be gained from using certain broad band excitation signals as well.

Experimental Results with Simulated Structural Faults: Over the past several years, several researchers and I have conducted some simple experiments to test our ideas regarding the use of modal parameters as a diagnostic for detecting, locating and quantifying structural faults. Our findings are reviewed here. More details are given in references [1] through [5].

Detecting Removal of a Bolt: An aluminum plate with a rib stiffener bolted along its centerline was tested before and after a bolt was removed from it. A diagram of the plate with rib stiffener is shown in Figure 5. (*See* Reference [1] for more details on this test.)

Figure 3 shows the measured modal frequencies for the first seven modes of the structure before and after the center bolt was removed. Shifts in all of the modal frequencies due to this simulated structural fault are clearly indicated.

Figure 4 shows the Modal Assurance Criterion (MAC) values between the mode shapes from before and after the bolt removal. (The MAC calculation is equivalent to the *Dot* product (or *Scalar* product) between two modal vectors.). If the mode shapes didn't change, the MAC matrix would have ones (1's) on its diagonal. It is clear from Figure 4 that substantial changes occurred in the mode shapes of modes 4 & 5 due to the center bolt removal. *It is also clear that the mode shapes of the other 5 modes changed very little.*

Figure 5 shows the mode shapes of modes 4 & 5 before and after the bolt removal. Even though they look the same at many points, the drop in MAC values indicates that they have changed.

	With Bolt	Without Bolt	
MODE	FREQ (Hz)	FREQ (Hz)	DIFFERENCE (Hz)
1	106.687	105.635	-1.052
2	190.636	190.186	-0.450
3	247.650	242.994	-4.656
4	259.222	254.200	-5.022
5	261.955	260.137	-1.818
6	470.489	466.324	-4.165
7	494.810	484.482	-10.328

Figure 3. Modal Frequencies Before and After Center Bolt Removal.

		With Bolt						
	MODE	1	2	3	4	5	6	7
Without Bolt	1	.98	0.00	.01	0.00	.01	.06	0.00
	2	0.00	.97	0.00	0.00	0.00	0.00	.01
	3	0.00	0.00	.96	.01	.03	.02	0.00
	4	0.00	.02	0.00	.62	.33	0.00	.01
	5	0.00	.01	.01	.35	.54	0.00	.01
	6	.06	0.00	0.00	0.00	0.00	.95	0.00
	7	.01	0.00	0.00	.02	.01	.01	.94

Figure 4. Mode Shape MAC Values Before and After Center Bolt Removal.

42

Mode: 4
Freq: 259.22 Hz   Damp: .12 %
Mode: 11
Freq: 254.28 Hz   Damp: .13 %

Mode: 5
Freq: 261.96 Hz   Damp: .89 %
Mode: 12
Freq: 260.14 Hz   Damp: .19 %
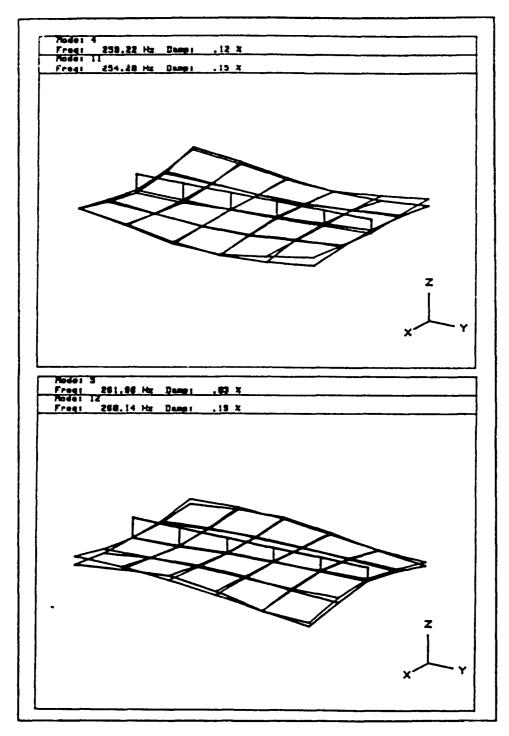
Figure 5. Mode Shapes of Modes 4 & 5 Before and After Center Bolt Removal.

Detecting Drilled Holes: To demonstrate the sensitivity of modal parameters to minute structural changes, several holes of different diameters were drilled in both an aluminum and a steel plate. (*See* Reference [5] for details.) Figure 6 shows the size of the plate and the holes, drawn to scale. The thickness of the aluminum plate was 10mm, and the thickness of the steel plate was 3mm.

FRF measurements were made on the plates before and after each of the holes was made in them. Five measurements were made for each case. Figure 7 shows a Modal Peaks Function for the Aluminum plate with no hole in it. (A Modal Peaks Function is the average of the *imaginary part squared* of the 5 FRFs.) There are about 40 modes in the frequency range of the FRFs.

Figure 8 shows expanded views of the Modal Peaks Functions in the frequency range of just two modes (1.92 kHz to 2.04 kHz). The three graphs superimpose the Peaks Function of the plate with no hole on the Peaks Function of the plate with three different sized holes; 2mm, 7mm, and 12mm.

The expanded views reveal that the 2 modes chosen clearly indicate the presence of the 12mm hole, by the frequency shift of the modes (Figure 8.C). Shifts of these two modes partially detect the 7mm hole (Figure 8.B), and don't visually indicate the presence of the 2mm hole (Figure 8.A).

Figures 9 and 10 show the same kind of results for the steel plate as those in Figures 7 and 8. Again, the expanded views of the Modal Peaks Functions reveal that the 12mm hole is easily detected, the 7mm hole is marginally detected, and the 2mm hole is not detected.
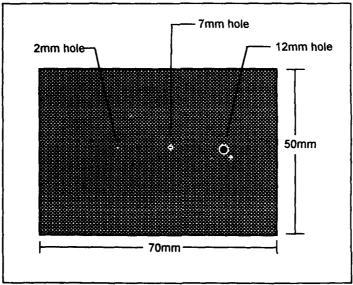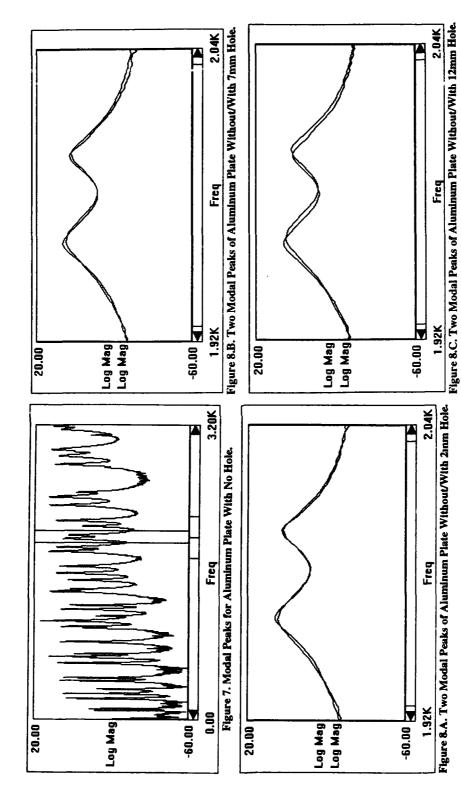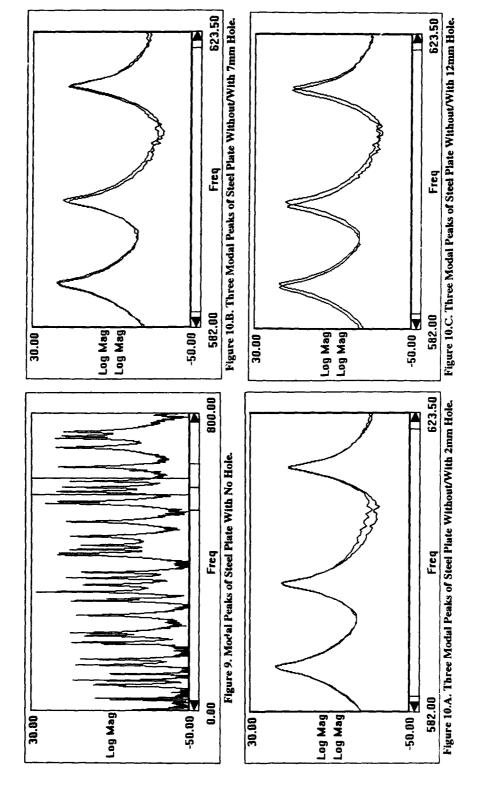
Figure 6. Plate Structure Showing Holes to Scale.

44

Figure 7. Modal Peaks for Aluminum Plate With No Hole.

Figure 8.A. Two Modal Peaks of Aluminum Plate Without/With 2mm Hole.

Figure 8.B. Two Modal Peaks of Aluminum Plate Without/With 7mm Hole.

Figure 8.C. Two Modal Peaks of Aluminum Plate Without/With 12mm Hole.

45

30.00

Log Mag

-50.00

0.00   Freq   800.00

Figure 9. Modal Peaks of Steel Plate With No Hole.

30.00

Log Mag
Log Mag

-50.00

582.00   Freq   623.50

Figure 10.A. Three Modal Peaks of Steel Plate Without/With 2mm Hole.

30.00

Log Mag
Log Mag

-50.00

582.00   Freq   623.50

Figure 10.B. Three Modal Peaks of Steel Plate Without/With 7mm Hole.

30.00

Log Mag
Log Mag

-50.00

582.00   Freq   623.50

Figure 10.C. Three Modal Peaks of Steel Plate Without/With 12mm Hole.

Locating a Saw Cut: In this experiment, an aluminum plate of dimensions (500mm x 190mm x 8mm) was tested both before and after a small saw cut was made in one of its edges. (*See* References [2] -[4] for details.) The location of the saw cut is shown in Figure 12. Fifty-five (55) FRF measurements were made on the "undamaged" plate using a uniform point grid on its surface. (Some of the test points are labelled in Figure 12.) The FRFs were then curve fit to yield the modal parameters for the first 10 modes of the plate.

The modal frequencies and mode shapes of the undamaged plate, plus the modal frequencies of the plate with the cut in it, were used to define a set of sensitivity equations. The sensitivity equations were then solved for the location of the *maximum negative stiffness change* on the plate. This iterative search process is depicted in Figure 11. Four iterations of the solution process are shown in Figure 12. The lines indicate the DOFs between which the maximum negative stiffness changes occurred. After the fourth iteration, the location of the maximum negative stiffness change coincided with the location of the Saw Cut.

Fault Detection: Fault detection is relatively straightforward if based only on modal frequency and damping changes, since changes in these parameters can be determined from practically any FRF measurement taken from a structure. The above experimental results demonstrate that modal parameter changes are indeed a sensitive indicator of structural faults. The implementation of an on-line structural monitoring system that uses modal parameter changes as a means of *only detecting* structure faults is definitely feasible using current day technology. Locating and quantifying a fault is a much more complex matter, however.
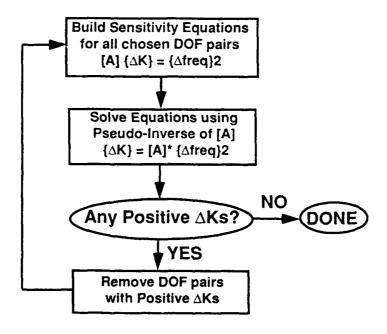
Figure 11. Maximum Negative Stiffness Search Method.

47
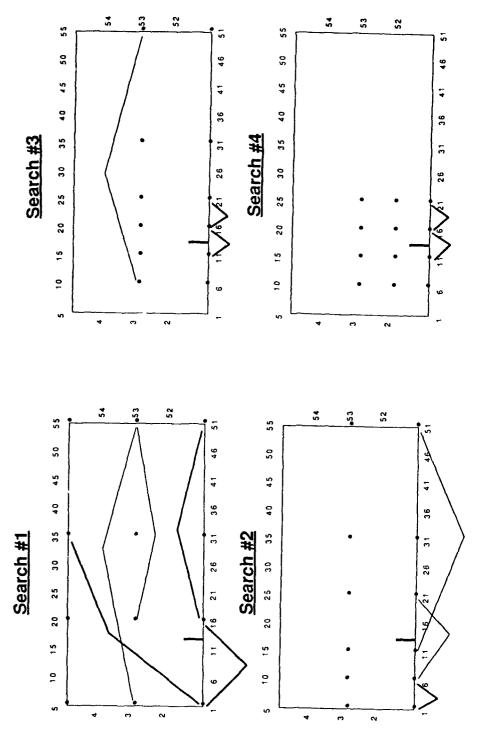
Figure 12. Four Iterations of the Search for the Saw Cut Location.

Fault Location and Quantification: To locate and quantify a fault, an accurate set of modal parameters (frequencies, damping, and mode shapes) for *all of the dominant modes* of the undamaged structure must be known beforehand. The number of DOFs of the mode shapes determines the degree of the *spatial resolution* that is possible for locating the fault.

To reliably locate faults, modal data for the higher frequency local modes is required. If the fault is very localized, then the local modes with non-zero mode shapes in the vicinity of the fault will be affected most.

The method used to locate a saw cut in the experiment described above assumed that the fault was predominantly a stiffness loss that could be located by finding the region of *maximum negative stiffness change* on the structure. This approach relies on the solution of an underdetermined set of equations (many more unknowns than equations) to find the maximum negative stiffness changes. Further development of this technique is needed to insure that it yields reliable, repeatable results.

Neural Networks: In any practical on-line monitoring system, a pattern recognition scheme will be needed to decipher the complex pattern of modal parameter changes that occurs due to a fault. Neural networks are proving to be an effective tool for pattern recognition in a variety of applications.

Neural networks were developed to mimic the pattern recognition capabilities of the human brain. Recently, they have been successfully implemented in Optical Character Recognition (OCR) software with a success rate in the high 90 percents, far exceeding previously tried statistical methods.

SDM and Neural Network Training: A key requirement of the use of a neural network is that it be "trained" beforehand. In this application, training the network would involve feeding it sets of modal parameter changes along with the mass, stiffness, and damping changes that caused them. The neural network, in turn, evolves (computes) a set of internal weights that allow it to predict the mass, stiffness, and damping changes that caused the modal parameter changes.

The SDM algorithm (*See* Reference [6]) can compute the new modal parameters due to mass, stiffness, or damping modifications very rapidly, using only experimental modal data to describe the dynamics of the undamaged structure. SDM can therefore be used to generate the numerous sets of data (mass, stiffness, damping changes paired with modal parameter changes) required to train a neural network. Random mass, stiffness, and damping changes could be fed to SDM to generate the required modal parameter changes.

Once a network has been trained for a particular structure, it can be implemented in an on-line monitoring system that will predict the *location* and *severity* of any fault that causes changes in the structure's measured modal parameters.

49

Conclusions: To answer the original question in the title of this paper, several experimental results were presented which demonstrated that modes can be used as a diagnostic for structural fault detection. The more complex issue of fault location and quantification was also discussed, and some suggestions for further research were given.

All of the tools necessary to implement an accurate and sensitive on-line structural health monitoring system are available in current-day modal testing technology. Improvements in the speed and accuracy of FFT-based signal analyzers, frequency domain modal parameter estimation, and recent successes in the application of neural networks to real world pattern recognition problems make the implementation of an on-line structural fault detection system a practical reality.

References

[1] Wolff, T. & Richardson, M. H. **"Fault Detection in Structures from Changes in Their Modal Parameters"** 7th MAC Proceedings, Las Vegas, Nevada, January 30, 1989.

[2] Mannan, M.A. and Richardsion, M.H. **"Detection and Location of Structural Cracks using FRF Measurements"** 8th IMAC Proceedings, Kissimmee Florida, February, 1990.

[3] Mannan, M.A. and Richardsion, M.H. **"Determination of Modal Sensitivity Functions for Location of Structural Faults"** 9th IMAC Proceedings, Florance, Italy, April, 1991.

[4] Richardson, M. and Mannan, M.A. **"Remote Detection and Location of Structural Faults using Modal Parameters"** 10th IMAC Proceedings, San Diego, California, February, 1992.

[5] Richardson, M. and Mannan, M.A. **"Correlating Minute Structural Faults with Changes in Modal Parameters"** 11th IMAC Proceedings, Kissimmee, Florida, February, 1993.

[6] Wallack, P., Skoog, P., and Richardson, M. H. **"Simultaneous Structural Dynamics Modification (S²DM)"** 6th IMAC Proceedings, Kissimmee, Florida, February, 1988.

[7] Richardson, M.H. and McHargue, P.L. **"Operating Deflection Shapes from Time Versus Frequency Domain Measurements"** 11th IMAC Proceedings, Kissimmee, Florida, February, 1993.

[8] Caudill, M. and Butler, C. **"Understanding Neural Networks, Volumes I and II"**, MIT Press, Cambridge, Mass. 1992.

ADVANCES IN THE DESIGN AND ASSESSMENT OF RELIABLE ELECTRONICS USING CONCURRENT ENGINEERING

Michael Pecht, Ph.D, PE
CALCE Electronics Packaging Research Center
University of Maryland, College Park, MD 20742

Abstract: Only with the advent of concurrent engineering concepts, has the electronics community begun to re-evaluate how they address reliability. The results of this self-study have been astonishing, as the overwhelming conclusion is that many of the reliability concepts employed to design reliable electronics have negatively impacted the success of many electronic systems. This paper discusses some of the history behind the problems and overviews some new directions in concurrent engineering.

Key Words: Concurrent engineering; design for reliability; electronics; physics-of-failure.

Introduction: Even before Wohl investigated and conducted tests on the fatigue of railroad axles, in the 1860s, have mechanical and civil engineers understood the role of reliability, from a physics of failure and root cause perspective. Today, the philosophy of physics of failure is commonplace in the design and assessment of most mechanical devices and structures. However, in the field of electronics, the physics of failure approach has not been widely applied.

The electrical reliability engineering discipline had it's beginning in the 1950s. At that time, the military was a dominant and largely dissatisfied, customer of electronic devices [Coppola, 1984]. As an example, during this time, the Navy was supplying a million replacement parts a year to support 160,000 pieces of equipment. Concern by the military led to the development of numerous standards and handbooks to address reliability tasks and methods. Unfortunately, because the military was not involved directly in the design or manufacture of electronics, their emphasis was on simple calculations which did not require expertise in electronics, or knowledge of the materials and packaging architectures from which the electronics were comprised. As an example of this philosophy, in the first handbook on electronics reliability, MIL-HDBK-217A, published December 1, 1965 under the Preparing Activity of the Navy, the reliability for all monolithic integrated circuits, was given by a single point failure rate of 0.4 failures per million hours, regardless of the stresses, the materials or the architecture. This single valued failure rate was illustrative of a philosophy that accuracy was less of a concern than ease of use, consistency and standardization; a philosophy which remains intact today by some military and government organizations.

In July 1973, RCA proposed a new set of prediction models for microcircuits, based on previous work by the Boeing Aircraft Company. The models were more difficult to use because they required familiarity with device fabrication techniques, materials and geometries, and utilized a lognormal distribution. These models were greatly simplified in-house by the Air Force at Rome Laboratories (formerly RADC) by removing most of the parameters and assuming an exponential failure distribution. The model simplicity and exponential distribution assumption still remain in MIL-HDBK-217 today, in spite of overwhelming evidence that distributions such as log-normal or Weibull are much more appropriate.

The advent of rapid changes in electronics and more complex microelectronic devices, pushed the application of MIL-HDBK-217 totally beyond reason. A good example was the MIL-HDBK-217 calculation of 13 seconds as the mean time between failure for common, and highly reliable 256K memory devices. The Air Force Rome Laboratories could not keep pace with the accelerating and ever changing technology base, in spite of updates made on the average of every seven years. Furthermore, because of the philosophy adapted by the Air Force Rome Laboratories, even in the latest MIL-HDBK-

217 update, the major recommendations by the contract companies, IIT Research Institute, Honeywell SSED, Westinghouse, and University of Maryland, were ignored.

Reliability in Concurrent Engineering: While not the sole reliability tool, the prediction of reliability is an integral part of the design, manufacture and operation of a product. An overview of the various reliability tasks and the reliability prediction input is given below.

Allocation. Allocation entails the assignment of reliability goals to the equipment and the subsequent assimilation of reliability to sub-systems, assemblies and parts. That is, commencing with an overall goal for product reliability, allowable reliabilities are apportioned. The reliability goal must be based on the expected life cycle design usage environment and the product mission profile.

System Architecture and Device Specification. As the physical design begins, reliability analysis can affect the system architecture and part selection, although functional and performance characteristics play the dominant role. Individual components must not be considered to be the only, or necessarily the major, source of failures. Interconnections and structures must also be properly selected. Another aspect of system architecture is the use of redundant circuits. Redundancy may be deemed necessary for mission completion when the estimates on reliability indicate improbable success or unacceptable risk.

Stress Analysis. Given the system architecture and parts, reliability models are used to assess the influence of the magnitude and duration of the stresses on the reliability of the parts and systems, so that stress and environment controlling systems (i.e., vibration, and cooling systems) and derating techniques can be implemented. Temperature, humidity, electrical fields, vibration and radiation are major stress variables affecting reliability.

Derating. Derating is based on the concept that operating electrical, thermal-mechanical and chemical stresses accelerate failures in a predictable manner, which if controlled, will improve reliability. For electronics, typical derating parameters include current, voltage, power, fanout, frequency, and operating (i.e. junction) temperature. Using the mathematical expressions of reliability prediction, one can often derive a derate schedule. Such schedules must be based on the dominant failure mechanisms for the particular electronics and must include interconnects and device interactions, as well as the devices themselves.

Environmental Controls. There are various ways in which both the operating and environmental stresses can be controlled to improve reliability. Methods can be applie ·o keep harmful stresses (i.e. high temperatures, high shock loads, high humidity, high radiation etc.) ·· ay fro.n sensitive devices and structures, or to manage the system environment to obtain controlled stress conditions. Lowering the harmful stresses is often a first choice of designers for reliability improvement. However, the cost and complexity of lowered stresses must be balanced against the cost and complexity of electronic complications to improve reliability by improved architectures and parts.

Stress Screening. Screening is the process by which defective parts, resulting from improper or out of control manufacture and assembly processes are detected and eliminated from a production batch. The principle involves inducing latent defect failures only in a population of parts that has already "weak" parts without reducing the reliability in the population of "strong" parts. The assumption is that, through the application of short-term stresses, the weak population can be discovered and eliminated, leaving a highly reliable population. Stress screening and burn-in (i.e. high temperature screen) methods are often based on reliability prediction models, and the acceleration stress levels are often derived from the models for the potential failure mechanisms associated with potential problems in quality.

Failure Modes, Effects and Criticality Analysis (FMECA). FMECA is a method to assess the inter-operability of the parts, sub-assemblies, assemblies and sub-systems comprising the system. The objectives are to: determine the effect of failures on system operation; identify the failures critical to

52

operational success and personnel safety; and rank each potential failure according to the effects on other portions of the system, the probability of the failure occurring, and the criticality of the failure mode. Reliability predictions are often used to determine the probability of failure for each potential failure modes of each element in the system.

Maintainability and Logistics. Maintainability assessment often uses failure rate data from reliability prediction models to determine a mean time to repair (MTTR) from element times to repair. The MTTR and metrics associated with acquisition, personnel, business and other issues are then used, along with reliability predictions, to calculate logistics parameters such as availability and supportability. It is critical that the design team realize that errors in the reliability predictions can be multiplied many times in the calculation of logistics metrics.

Certification. This is the culmination of the product development process, where it is agreed that the product is ready to be introduced to the market, having met or exceeded marketing, contractual, regulatory, or other goals for performance. Where reliability is an item affecting this final decision, many if not all of the foregoing reliability tasks will be involved.

Warranty. The expectations of reliability often affect the warranty terms. In some cases, suppliers may be required only to meet contractual goals without incentive for, or interest in, continued reliability improvement. That is, the concept of "attainable maximum" often provides an easily achieved cap on expectations. There are many other warranty arrangements, often intended to encourage suppliers to treat product reliability seriously. For example, the desired reliability goal bears economic considerations that affect life cycle cost. Those costs are usually included in the fundamental economic analysis to determine economic feasibility of the total program, and in some cases can be an important item in total costs of ownership.

Failure Diagnosis and Corrective Actions. Failure diagnosis and corrective actions may be involved as part of a continuous product improvement program. When the goal is only to meet warranty requirements, there is seldom any interest in further diagnosis and corrective action after meeting goals. In such an instance, reliability prediction can provide the basis for a hindrance to continued improvements in reliability. Reliability growth is associated with the continuous improvement in product reliability. However, once again, the calculated reliability should not necessarily be considered to be the maximum achievable reliability.

Cost Effectiveness. Many variables affect cost effectiveness. Cost, weight, volume, dependability, and a myriad of other factors can all have a role, and thus cost effectiveness studies can be quite complex. When reliability is a major element as is the case with aviation equipment, dollar cost can be less significant than other factors such as weight, volume, and power consumption. All costs must be defensible in terms of product value.

A New Reliability Paradigm: To address the role of reliability in the concurrent engineering process, a new reliability paradigm has been adapted. This paradigm focuses on understanding potential failure mechanisms and the root cause of failure in order to provide reliability goals expressed in terms of a failure free period, rather than providing un-verifiable statistical means in which the tails of the distribution have not been investigated. In terms of manufacturing, the qualified manufacturing process replaces qualification of parts as a means to confront potential defects when they first arise. Finally, the concurrent engineering approach replaces the sequential approach system engineering. A comparison of the new and old paradigm is given below. An example will then be given to overview the new paradigm.

OLD	NEW
• Reliability predictions based on field performance data and statistically-based constant failure rate models.	• Reliability predictions based on physical failure phenomena, each with its own distribution in time. Data obtained from structured testing on recent technologies. Stress and acceleration factors based on physics of failure.
• Reliability goals expressed by MTBF.	• Reliability goals expressed as lifetimes.
• Part qualification.	• Process qualification.
• Reliability at any cost guaranteed by mandated methods.	• Cost effective reliability guaranteed by robust design and manufacturing processes.
• Sequential life cycle engineering	• Concurrent engineering

Example: The Computer Aided Design of Microelectronic Packages (CADMP) design and assessment software environment is an example of how reliability tasks can be implemented into the concurrent engineering environment. CADMP is a set of integrated software programs for design and assessment of IC, hybrid, and multichip module packages. The benefits of using this software include scientific consideration of reliability during the design phase; evaluation of new materials, structures, and technologies; assessment of packages designed by different manufacturers; ability to develop science-based tests, screens, and derating methods; and cost-effectiveness product development achieved by investigating trade-offs of various design options.

The potential users of this software include:

• Designers, manufacturers, and testers of IC, hybrid, and MCM packages
• Companies which use packaged ICs, hybrids, and MCMs for circuit cards
• Agencies which assess, evaluate, test, or define specifications and requirements for IC, hybrid, and MCM packages, or circuit cards and equipments with these packages.

The CADMP software is developed based on the physics-of-failure principles. The physics-of-failure approach to reliability utilizes the knowledge of failure mechanisms and the root causes of failures to address product failures through robust design and manufacturing practices. In the physics-of-failure approach, average time to failure based on stresses, material properties, geometry, environmental and usage conditions is determined for each potential failure mechanism and associated failure site. Potential failure mechanisms and associated failure sites can be ranked and weak links in the package can be identified, and test and derating methods developed.

The CADMP software considers the total electronic package, as well as package elements such as substrate, attachment, interconnect, case, lead, lead seal, and lid and lid seal. Potential failure mechanisms and failure sites, associated with the package elements and their interactions, are used to guide the selection of the materials and design of the package architecture.

54

The CADMP software consists of the following tools:

- Mission profile. The mission profile program provides a means to specify a series of test conditions, as well as storage and operating environments and their durations that the product is expected to experience. A library of common default environments for test, storage, usage, and transportation are provided in an environment database.

- Constraints management. The constraints management program provides a means to specify constraints and defaults for the package. The number of package I/Os, lead pitch and style, and package height and area, and board material are examples of constraints that can be specified.

- Geometry design, material selection, and die placement. The geometry design and material selection program provide the tools to design an electronic package from a selection of package templates. Templates include dual-inline package (DIP), quad flatpack (QFP), pin grid array (PGA), single inline package (SIP), small outline package (SOP), land grid array (LGA), and leaded and leadless chip carriers. This program also provides an aid for selection of package type and mounting technology, interconnect technology, and substrate technology, subject to any specified constraints.

- Stress analysis. The stress analysis program includes thermal and vibration analysis. These analysis provide temperature and mechanical stress-strain information to the failure analysis program and shows the affects of test and operation on reliability.

- Life prediction, testing, screening and derating based on potential failure mechanisms and associated failure sites. The goal of these programs is to assess packages designed by the CADMP software as well as those designed by different manufactures, provided that material and geometry information is available for each package element. These programs enable design and evaluation of new materials, structures, and technologies and design of science-based tests, screens, and derating methods.

Conclusions: The use of reliability predictions in the design and operation of electronic equipment has been an evolutionary and very controversial process. While it is generally believed that prediction methods should be used to aid in product design, assessment and support, often the integrity and auditability of the methods and models have been found to be questionable. In fact, the handbooks which have been developed to model reliability often do not accurately predict field failures, cannot be used for comparative purposes, and present misleading trends and relations.

With the advent of concurrent engineering, the need to introduce reliability within the design process became visible. A number of government programs are now utilizing the new reliability paradigm and developing tools and techniques. The economic benefits are already visible, as seen by the a number of projects sponsored by the Air Force Wright Patterson under the Avionics Integrity Program (AVIP) and by the Army AMSAA.

MFPG WORKING GROUP
BUSINESS CASE FOR MECHANICAL FAILURE PREVENTION

K. A. Krewer,
NAVSEA 92L, Naval Sea Systems Command
2531 National Center Bldg. 3,
Washington, DC 20362-5160

Abstract: This paper identifies a workable program (for discussion at the 46th MFPG conference) for a maintenance and monitoring system that prevents or predicts mechanical failures. It offers a list of operating parameters that represent the internal condition of equipment, a system and component selection process and a means of assessing the financial desirability of monitoring.

Key Words: PREDICTIVE MAINTENANCE, COST ANALYSIS,
        MONITORING PARAMETERS

Introduction: The idea of monituring the performance of mechanical equipment by replacing components when failure is imminent is nonintrinsic to most of our lives. We work on our cars when we hear a noise or a rattle indicative of a problem. We do preventive maintenance at a fixed interval but frequently postpone it until a problem is evident or the complexity / cost of the maintenance is justified. Each of us in our own way completes a business case assessment. We work on the things that we have to as they emerge and assess the benefits of maintenance to the cost. The concept of having predetermined intervals where we listen or check operation is important for large complex systems or where corrective action is not evident, such as checking the oil level in our cars and changing the oil every 3000 miles or changing belts at 10,000 miles.

Navy Ships, Nuclear Power Plants, and the Alaskan Pipeline are examples of systems that require major efforts to be maintained operational. Unlike our cars operators don't "hear" the problems until they are upon them and the cost of stopping and fixing the problem are prohibitive. There needs to be a way of determining the need to perform corrective or preventive maintenance and avoid failure. If you are too conservative in your prediction you spend money needlessly, if you are too liberal you have a failure and a premature shut down.

What is involved in a predictive maintenance program? There are many companies that specialize in developing

programs specifically for predictive or condition based maintenance. Many are very good and can be of great assistance in providing information of how monitoring was done successfully in the past. As a designer or operator you understand the interrelation of the systems that produce your product. Therefore, you are in the best position to understand the effects of loosing operational time of a component, and the effect this has on other equipment. The basis of any mechanical failure prevention is to stop and complete maintenance before degradation impacts performance acceptably. Figure I, is a flow diagram which could be used to assess the advisability of monitoring the condition of a system or component. Figure II is a flow diagram for selecting initial systems and components for monitoring consideration. Those systems that are primary to safe and reliable operation are prime candidates for monitoring. The next step is to determine if there are parameters that represent the condition, performance, and operability about the system. With out indicative relationships, failure may not be evident. With improvements in technology and trending we are better able to see patterns and predict in equipment performance. Figure III is a sample list of parameters used to monitor the performance condition of a component.

The cost of implementing and maintaining a program includes both short term and long term operating expenses. Frequently with the initiation of monitoring, hidden problems surface and maintenance costs increase. Initial costs of developing a program and acquiring test equipment can be done small scale or with an automated approach depending on resources available. Monitoring can be an iterative process where different approaches are taken to identify the most advantageous and easiest way of collecting and analyzing data. A balance of importance to ease must be developed so the greatest benefit per dollar may be achieved. Figure IV, is a life cyle representation of a performance monitoring program.

Non-tangible benefits and lost opportunities are weighted to fully evaluate the cost. Monitoring is not the answer for all problems and may not bring the return for investment that is necessary to secure funds. Reviewing Figure V will help you better understand the application and considerations involved with developing a monitoring program and allow you to select the type and depth of monitoring which is most cost effective for you.

The business case is a mechanism of establishing a way to evaluate a change process. Standard Cost Benefit

principals applied with an analytical approach
successfully compare alternative programs needing
funding. The business case has a distinct benefit in
that it documents the thoughts and relative worth of
ideas at the time of decision and provides a historical
reference. The relative value of non-tangible costs and
assets give us a better understanding of specific
considerations at the time of decision. Variations in
value weighing can sway the results of a cost
evaluation. A mechanism of revisiting and refining the
cost benefit model is incorporated in an effective
business case. The old saying that "the study is not
done until all the data is in" holds true. After a
program is implemented and operational, it is best to
reassess and see if you achieved your return for
investment estimate. Figure VI is a modification to a
Balance Scorecard.

Figure I
FLOW DIAGRAM FOR RECOMMENDED MONITORING

SYSTEM OR COMPONENT

SYSTEM/COMPONENT DEGRADATION ————————➤ YES
& FAILURE IS NOT EVIDENT

NO

IS SYSTEM/COMPONENT CRITICAL TO PROCESS FUNCTION ——➤ YES

NO

IS SYSTEM/COMPONENT CRITICAL TO SAFE OPERATION ———➤ YES

NO

WILL SYSTEM/COMPONENT DEGRADATION AND OR FAILURE ——➤ YES
CREATE A SAFETY HAZARD?

NO

WILL SYSTEM/COMPONENT UNRELIABILITY REQUIRE HIGH ——➤ YES
LOGISTIC RESOURCES EXPENDITURE

NO

DOES SYSTEM/COMPONENT LACK REDUNDANCY ———➤ YES

NO

YES ——— IS SYSTEM/COMPONENT EASILY MAINTAINED ———➤ NO
OR REPAIRED?

NO   WILL MONITORING ◄——┘
DETECT FAILURES & BE
PRACTICAL & EFFECTIVE

YES

SYSTEM /COMPONENT NOT
RECOMMENDED FOR MONITORING

SYSTEM/COMPONENT
RECOMMENDED FOR MONITORING

59

FIGURE II
INITIAL SYSTEM AND COMPONENT
SELECTION PROCESS

```
          COMPONENTS WITHIN THE SYSTEM ARE ROUTINELY
   ┌──────────────── CHANGED OUT BASED ON TIME
YES                          │
   │                         │ NO
   │                         ▼
   │              SYSTEM REFURBISHMENT
YES ◄─────────── REQUIRES SPECIAL MAINTENANCE
   │    (e.g. plant shut down, long lead time for parts)
   │                         │
   │                         │ NO
   │                         ▼
YES ◄───── SPECIAL MAINTENANCE PERSONNEL MUST BE
   │         CONTRACTED FOR REPAIR OR REFURBISHMENT
   │                         │
   │                         │ NO
   │                         ▼
   │   MAINTENANCE OF COMPONENTS IS MINOR (e.g., clean,
   │              inspect, lubricate, test,
   │ NO ◄──────── calibrate, periodically cycle) ──────►YES
   │
   │
   ▼      SYSTEM OR COMPONENT IS SAFETY, RELIABILITY       ▼
YES            OR CRITICAL TO PRODUCTION PROCESS ──────►NO

      INCLUDE IN MECHANICAL              DO NOT INCLUDE IN
   FAILURE PREVENTION PROGRAM            MECHANICAL FAILURE
                                         PREVENTION PROGRAM
```

FIGURE III
LIST OF MONITORING PARAMETERS

The following is a list of parameters to select from.
They may be measured either externally or by installed
in line sensors.

Vibration
Thermal Imaging
Electrical Voltage and Current Signature
RPM
Power
Electronic Emanations
Flow
Temperature
Insulation Resistance
Leakage Rate
Valve Timing
Oil Analysis (spectra analysis, UV transmission)
Eddy Current Transmission
Torsion
Pressure (operating and other)
Volumetric Flow
Flow Regime
Combustion Products
etc.

60

FIGURE IV
LIFE CYCLE OF PERFORMANCE MONITORING

Collect Monitoring and Maintenance Data

Revise Criteria or
Technology Insertion

Assess Data

Evaluate
Design & as Found Condition
Recommendations
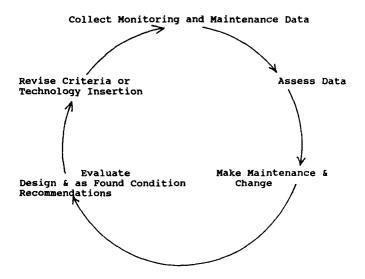
Make Maintenance &
Change

FIGURE V
MONITORING START UP FLOW DIAGRAM

System selection

Component selection

Failure Analysis & Reliability Modeling

Monitoring Parameters selection

Evaluation Criteria Development

Hardware purchase or Development

Equipment Modification

Technician Training

Analysis software and Hardware

Cost Estimate

Benefit Estimate

Business Case work Evaluation

61

FIGURE VI
SAMPLE BALANCE SCORE CARD

For each area a list of goals and measurement indicators are developed. The relative value is assessed and a weighting factor assigned. A value of 1 through 10 is assigned with the highest value assigned to the most important.

FINANCIAL PERSPECTIVE

A straight cost to return is calculated. (Not taking into account non-tangible benefits.) The cost will include long term equipment purchased, operator training, and any equipment modifications necessary. Savings will include maintenance eliminated and 50% of the cost of maintenance deferred for over 12 months.
Factors:
1) % return on investment required for management support,
2) Inflation rate of postponing investment,
3) Lost opportunity cost (predicted rate on investment from alternative plans),
4) Interest on money borrowed,
5) R&D money payback for new technology development.

CUSTOMER PERSPECTIVE

Consider if there is a cash benefit for having a monitoring system in place to the customer it should be noted here. Customers may be more willing to pay a certain percentage for higher reliability, casualty avoidance and improved risk assessment.
Factors:
1) Increased reliability or safety,
2) Ability to planned shortages or down time,
3) Reduced emergent work,
4) Bulk buying or work load smoothing,
5) Rapid & reliable trouble shooting,
6) Historic recording and trending,
7) Improved operational design,
8) Efficiency improvement of consumed fuels.

INTERNAL PERSPECTIVE

Consider the improvements in management of the organization by looking at factors that increase the value of individuals.

Factors:
1) Improved training of operators to collect and trend data,
2) Maintenance of equipment used for monitoring, internally calibrated and repaired,

62

3) Development of engineering assessment of data or computerized expert system,
4) Development of procedures and time tables to implement monitoring,
5) Better understanding of systems approach to operations.

INNOVATIVE PERSPECTIVE

Consider if there is a benefit to the organization or other programs for developing technological improvements for monitoring.
Factors:
1) Development of self diagnostics,
2) Technology investment and product.

CONCURRENT ENGINEERING
FOR
MECHANICAL SYSTEMS

CoChairmen: Victor K. Champagne
              U.S. Army Research Laboratory

              Ruben A. Lebron, Jr.
              Naval Air Warfare Center

MECHANICAL AutoTEST:
A CONCURRENT ENGINEERING APPROACH TO INHERENT TESTABILITY ASSESSMENT OF HYDRO-MECHANICAL SUBSYSTEM DESIGNS

Douglas K. Bransford
Donald L. Gartner
Steven A. Wegener
McDonnell Douglas Corporation
St. Louis, Mo. 63166

Abstract: Concurrent engineering alters the traditional approach to design by emphasizing equipment characteristics such as reliability, maintainability, supportability, and testability in parallel with performance characteristics. This approach provides for incorporation of the diagnostic capability as an integral part of the system design.

The objectives of testability analysis are to introduce testability considerations early in the design effort and provide an assessment of the diagnostic capabilities of the design. Testability analysis utilizing dependency, simulation, or functional modeling can be employed to assess diagnostic capability and the impact of adding, removing, or relocating test points. Each modeling approach has its own advantages and disadvantages. McDonnell Douglas Aerospace (MDA) has developed a computer aided engineering (CAE) tool known as the Automated Testability Expert System Tool (AutoTEST) which utilizes functional modeling and object-oriented classification techniques as a basis for testability analysis of electronic systems. A version of AutoTEST is currently in development for analysis of the inherent testability of hydro-mechanical systems.

This paper describes the role of a CAE testability analysis tool (i.e. AutoTEST) in the concurrent engineering approach to hydro-mechanical design, the basic principles of the mechanical system version of AutoTEST, and how this approach to testability analysis improves upon the analysis provided by tools employing other modeling techniques.

Key Words: AutoTEST; concurrent engineering; dependency modeling; figures of merit (FOMs); simulation modeling; testability analysis

Introduction: The need for testable aircraft systems and equipment has long been recognized. The issue of how to ensure testable electronic designs was addressed by the creation of MIL-STD-2165, Testability Program for Electronic Systems and Equipment. The testability of non-electronic designs has subsequently been addressed by the creation of MIL-STD-2165A, Testability Program for Systems and Equipment.

The traditional approach to hydro-mechanical subsystem design has failed to provide formal testability analysis early in the design process. Emphasis has instead been placed on the development of performance capabilities, with the associated size/weight/power/cost/ reliability constraints. Consideration of other design characteristics, including testability, is often delayed until the configuration baseline has been established. In this situation, changes which might improve inherent testability may not be incorporated due to cost, weight, and/or schedule impact. The penalties for hardware changes at this point in design have led to a reactive approach to diagnostic development. Fault detection and isolation methods are developed using whatever test points are provided. A proactive approach is needed to improve the inherent testability of system designs. It is imperative that testability analyses be performed during the conceptual and preliminary phases of design, while the design can be influenced with a minimum impact to cost and schedule constraints.

Concurrent engineering alters the traditional approach to design by emphasizing equipment characteristics such as reliability, maintainability, supportability, and testability in parallel with equipment performance. This approach provides for incorporation of the diagnostic capability as an integral part of the system design. It encourages consideration of design characteristics which impact these

capabilities early in design, thus enhancing the feasibility of incorporating such considerations. The National Security Industrial Association's (NSIA) Automatic Testing Committee (ATC) DoD/Industry Diagnostics and Testing Project offers the conclusion that "within the concurrent engineering design approach lies the most promising solution for fielding a satisfactory weapon system diagnostic capability [1]." When support is considered to have an importance equal to performance, testability can appropriately be addressed concurrently with other system requirements.

Hydro-Mechanical Testability Analysis Requirements: The requirement to perform a formal testability analysis of hydro-mechanical designs in military aircraft is a recent development. A major testability need in aircraft hydro-mechanical systems is an improvement in the capability to consistently isolate to the correct line replaceable unit (LRU) at the Organizational level of maintenance. The detection and isolation of malfunctions in hydro-mechanical equipment at the Organizational level is performed utilizing embedded diagnostics and/or flight line test sets. Historically, embedded diagnostics have not fully achieved the desired results. Testability analysis methods developed for electronic equipment generally focus on compatibility with automated test equipment (ATE) at the Depot level of maintenance and with aircraft embedded diagnostics or flight line test sets at the Organizational level. The complexity of the equipment and the testing required to verify its operation necessitates the use of ATE for off-aircraft testing of these devices. Except for a few isolated cases, off-aircraft testing of hydro-mechanical aircraft components has not traditionally utilized ATE in either the USAF or USN maintenance community. Due to this situation, the primary need for a testability analysis tool for hydro-mechanical designs exists at the system level rather than the component level.

The need to perform testability analysis exists during several design phases. The amount of detailed design information available will vary during each of these phases. Testability analysis during the conceptual and preliminary phases of design will be used to assess the impact of test point (or sensor) placement on system testability. This optimization will require iterative analyses to provide quantitative measures which can be used to assess a design's inherent testability. These measures are generally reported as a standardized set of figures of merit (FOMs) derived from customer specifications. The testability analysis method selected must be flexible enough to use whatever design details are available at any phase of the design. It must allow iterative analyses to be performed within the constraints of the design schedule.

Testability Analysis Methods: The capability to verify acceptable component operation and to detect and isolate faults is limited by the availability of, and access to, signal information in the system. Tests are developed to detect and isolate those faults envisioned by the test designer, using the test points provided in the equipment design. Testability analysis utilizing dependency, simulation, or functional modeling can assess the impact of including, excluding, or relocating test points. Several CAE tools have been developed to conduct testability analysis using dependency modeling techniques. Simulation models have also been used to determine fault detection and isolation capabilities. Additionally, MDA has developed a CAE tool known as the Automated Testability Expert System Tool (AutoTEST) which utilizes functional modeling and object-oriented classification techniques as a basis for testability analysis of electronic systems. A version of AutoTEST specifically for analysis of hydro-mechanical systems is in development.

Common Modeling Requirements: A critical aspect of any automated testability analysis tool is the way in which components are modeled. Essentially, a model based analysis determines a system's behavior from its components, their individual behaviors, and the interconnections between them. The capabilities of several modeling techniques to fulfill various testability analysis tasks have been demonstrated. Each of these modeling approaches has its own advantages and disadvantages. Certain common requirements exist for all modeling methods. These requirements include a description of the behavior of each component and a topological description of the design, to identify

how components are interconnected. The interconnection information is relatively consistent among each of the modeling methods; however, as will be discussed later, the contents and level of information required to describe component behavior differs depending upon the method of modeling employed and the level of accuracy desired. The topological description requires identification of the components, their interconnection, module ports (system inputs and outputs), and the allowed direction of information flow (input, output, or bidirectional) at each component port/pin. When creating a model of a device for a specific purpose, such as testability analysis, only the level of detail necessary to describe the behavior of interest need be represented in the model.

A desirable feature in any CAE tool is the capability to automate as much data entry as possible. System schematics and block diagrams created by computer aided design (CAD) tools provide an electronic data source for the topological description of the design. The availability of an interface which will allow transfer of this information from the designer's CAD tool into the testability analysis tool is an important feature in the selection of a testability analysis tool. In addition to topological descriptions, some standard data formats, such as the Product Data Exchange using STEP (PDES)/Standard for the Exchange of Product Data (STEP) and the VHSIC Hardware Description Language (VHDL), describe the functionality of each device at some level of detail. The capability to utilize this functional information to automatically create testability models is highly desirable.

Testability Analysis Using Dependency Modeling Techniques: Several testability analysis tools utilize dependency modeling to describe component behavior. Dependency models describe the component's output(s) in terms of its relationship with the events or inputs passed into the component and with the physical aspects of the component itself. An example is shown in Figure 1. A dependency is predicated on two fundamentals. The first is that passage of an output test implies that all associated component aspects are functioning normally and that all associated input tests also pass. The second is that failure of an output test implicates all associated component aspects and input tests as failure candidates.

In dependency modeling, both cause (input) and effect (output) events are usually called tests. Specific locations within the design are referred to as either nodes or test points. Components are usually referred to as items, and the physical portion of an item which relates to a particular function of that component is identified as an item aspect. The construction of a dependency model requires the ability to describe the cause and effect relationships involved in each test or event at each node. These relationships are shown for nodes A, B, and C of Figure 1. Item aspects may be defined either in terms of how an item works or how it fails. The amount of functional detail represented in a dependency model is reflected in the selection of the item aspects. The model may contain little more than topological information if the aspects are simple or may contain a great deal of functional and behavioral information if the aspects are detailed [2].

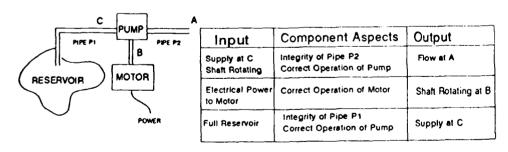| Input	Component Aspects	Output
Supply at C Shaft Rotating	Integrity of Pipe P2 Correct Operation of Pump	Flow at A
Electrical Power to Motor	Correct Operation of Motor	Shaft Rotating at B
Full Reservoir	Integrity of Pipe P1 Correct Operation of Pump	Supply at C

Figure 1

One concern with the automatic creation of dependency models from topological netlists is that these netlists provide only the most elementary aspect information. Models created in this manner require manual editing to reflect the level of detail needed to accurately influence testability analysis results. Other concerns associated with the use of purely topological dependency models for testability analysis are discussed below. These discussions assume that the aspect/test relationships utilized are selected to test for proper operation of device functions.

A concern with the use of dependency models for testability analysis is that the fault universe utilized in the analysis must be implied in the model. To provide a prediction of fault detection and isolation capabilities, one must be able to observe how faults are manifested in the design. In dependency modeling tools, faults are represented as the observance of a failed test. Device faults are implicitly represented by the test/aspect dependencies established in the models rather than explicitly represented as device faults. The ways that the device could fail and the tests which could be used to observe specific failure modes must be reflected in the combinations of aspects and tests selected to model the device. Even though aspects and tests may be described in terms of proper device operation, their "fail" or "no go" conditions represent device faults, and the need to define a fault universe within the model is emphasized during the creation of dependency models. The need to create models which emphasize device failure modes contradicts the natural inclination of a designer to model devices based on how they operate.

Another concern is that dependency models created from a purely topological description of the design are of insufficient depth to accurately identify ambiguity groups. As previously noted, dependency models may represent little more than topological information if the aspects are simple. Dependency models constructed from the topological description of a design generally represent first order dependencies in terms of a single test at each component output, a single aspect, and a single test at each component input. The output test is subsequently recognized as an input test to a downstream component. The single aspect used to describe the functionality of the component is typically designated as something similar to "the component functions properly". The rationale used to justify the adequacy of a dependency model containing a single aspect for use in testability analysis is that if the output test passes, the component is recognized as "good" and that if the test fails the component is recognized as a possible cause of the failure and is included in the appropriate ambiguity group. While this is true, consideration must also be given to the likelihood of whether or not a single test, or suite of tests, can be developed which is capable of verifying all operational modes of a device. An analysis based on a purely topological model can result in an inaccurate definition of ambiguity groups if the model does not contain a realistic representation of how various failure modes would be observed in the design. If separate aspects were developed to represent each function of a particular device, various aspects of the device could correctly be included in separate ambiguity groups. An analysis based on a purely topological model will result in the identification of worst case ambiguity group sizes and best case component involvement ratios (CIRs).

The need for early detailed design data to accurately place test points is also a concern. The test point placement recommendations provided by the most common dependency modeling tools are derived in part from the impact of the test points on ambiguity group size. Since analyses based on a purely topological model will result in the identification of worst case ambiguity group sizes, confidence is reduced in the test point recommendations based on these ambiguity groups. As discussed above, the representation of the fault universe for a dependency model must be implied by the tests and aspects contained in the model. This dependency on the observation of specific tests to determine how faults propagate through the model necessitates that the modeler possess a detailed understanding of the way the failure modes of the devices could be detected. The accuracy with which this can be represented in a dependency model is dependent on the amount of detailed design information available to the modeler. Because of the level of detailed design information required to create dependency models which

70

accurately reflect how faults propagate through the model, dependency models appear to be best suited for use in assessing whether or not the testability requirements allocated to the system design have been met.

Another consideration is that libraries containing dependency models of standard components or component types are not currently available. The amount of functional and behavioral information included in a dependency model for a particular design is left to the discretion of the individual creating the model.

To assess whether or not testability requirements have been met and how changes impact the inherent testability of a particular design, quantitative measures must be provided. These measures are generally reported as a standardized set of FOMs derived from customer specifications. The purpose of a FOM is to provide some measure of objectivity in the often subjective design process. One such set of FOMs is that generated by the Weapons System Testability Analyzer (WSTA) which is a dependency modeling tool developed as an element of the USN's Integrated Diagnostic Support System (IDSS). These FOMs include weighted fraction of faults detected (FFD), weighted fault isolation resolution to n components (FIR-N), component involvement ratio (CIR), weighted mean ambiguity group size (WMAGS), mean penalty to isolate (MPTI), and mean penalty to repair (MPTR).

Testability Analysis Using Simulation Models: Another approach to testability analysis involves the use of simulation models. Hydro-mechanical system designers utilize simulation models to calculate performance parameters (e.g. flow, pressure, temperature, etc.) at each node in the design. Detailed design information is required to calculate these parameters. This method of modeling includes much more detail than is required to infer inherent testability.

Most simulation tools and models used in hydro-mechanical system analysis were developed for unique applications and are considered proprietary by the design activities which utilize them. While simulation models can be used to indirectly assess the effects of design changes on system testability, this is not the purpose for which these tools were developed. The use of these tools in their current form for this purpose will require significant manual analysis in order to interpret the results and determine whether or not testability requirements have been met. To utilize these tools for testability analysis, the effects of component failures on system performance must be introduced through the component models. The equations contained in existing models represent the effects of a properly operating component on system performance. The representation of how a faulted component would affect system performance is currently undeveloped in the hydro-mechanical simulation models in use at MDA. The effects of some failure modes can be represented in the dimensional information input by the user as variables to these equations. The representation of other failure modes might require the development and verification of new models to calculate the effects of that particular fault on a component's performance.

Another consideration involving the use of simulation modeling tools for testability analysis is that these tools do not currently provide a means of generating FOMs. To produce FOMs using these tools a fault table must first be generated from the fault insertion results. Then an interface must be developed with a postprocessor, such as the WSTA FOM generator previously described, or else algorithms used to calculate similar FOMs must be added to the simulation modeling tools.

Testability Analysis Using AutoTEST: AutoTEST provides a method of modeling design information based on the principles of Model Based Reasoning (MBR) and object-oriented classification. The methods utilized allow the functional information required to perform a testability analysis to be modeled in a manner which overcomes some of the problems inherent in the use of dependency or simulation models for testability analysis.

AutoTEST development is part of an ongoing MDA effort directed at defining the Integrated Diagnostic (ID) process and the tools required for its efficient implementation. The general ID process and an outline of the tool requirements have been presented in two earlier

71

papers [3,4]. AutoTEST was initially developed to assess the testability of digital electronic circuits. The digital circuit analysis capabilities of AutoTEST are described in [5]. A derivative of the tool was subsequently developed to provide analog and system level testability analyses. These capabilities are described in [6]. The analysis techniques developed for the system level and analog tool are being utilized as the basis for the development of a version of AutoTEST for analysis of the inherent testability of hydro-mechanical systems.

Commonality between the analog electrical and hydro-mechanical versions of AutoTEST is critical because hydro-mechanical systems in fighter aircraft are hybrids. Many components and sensors utilized in these systems are electrically controlled or actuated, and the circuitry associated with the application of power to, or the return of signals from, these devices is considered part of the hydro-mechanical system. Purely mechanical systems are, for all practical purposes, nonexistent. To effectively assess the impact of design changes on testability at the system level, the testability analysis tool must be be usable for analysis of electrical, as well as hydraulic and pneumatic, devices.

Since testability analysis must be performed during the preliminary design phase or earlier to optimize test point selection, limited design data may be available. Component models require the representation of only those attributes which directly contribute to testability analysis. The modeling scheme developed for AutoTEST represents the flow of information through the component. The analysis techniques developed for the system level AutoTEST tool require that information flow at the component level be classified into four categories: signal, control, condition, and bias. Signal flow represents the primary information path(s) through a component. Control flow defines those inputs to a component which control or set its operational mode. Condition flow identifies those inputs which may modify or affect the signal flow, but are not required for the component's operation. Bias flow indicates those signals which must be connected to bias (a power source) for the component to operate properly.[6]

In the current applications of AutoTEST, the topological description of a design is obtained directly from a standard electronic data output format, the Electronic Design Interchange Format (EDIF) version 2.0, Level 0 flatfile netlist. This data format is a standard output for many of the CAE/CAD tools used to create electronic schematics. At present, hydro-mechanical system topological data can be input to AutoTEST by creating a block diagram representation using a tool such as OrCAD and generating an EDIF netlist of the system.

A topological description of the system is also required as an input to the CAE tools used for hydro-mechanical system simulation at MDA. Until recently, this description has been input from manually generated netlists. A newly developed CAE tool in use at MDA, the Design Knowledge Capture (DKC) portion of the Integrated Crew Chief's Associate (ICCA), provides the capability to generate a netlist for use in these simulation tools directly from the system schematic representation. Part of the concurrent engineering philosophy is that common data sources should be used for different design tools. To support this effort, the hydro-mechanical version of AutoTEST will include a data parser to allow it to accept the ICCA generated netlist. As CAD tools used to create hydro-mechanical schematics adopt standard data formats for netlist generation, a parser to utilize this data will be developed and incorporated into AutoTEST. The most likely standard data format for this application is PDES/STEP.

AutoTEST models are developed using an object-oriented hierarchy. This modeling method allows common attributes to be assigned to classes of components. When a component is initially modeled it is defined to be a member of a class and thus inherits all common attributes of that class. AutoTEST provides static model libraries of various component classes. Existing component models may be selected from the appropriate model library, or new component models may be created by the user and stored in a netlist knowledge base (KB). These newly created models will subsequently be reviewed by the AutoTEST development group, then added to the appropriate model library. Once

added to the library, these models can be utilized by other users. The classification of components based on their common functionality allows specific design for testability (DFT) rules to be applied to designs containing certain classes of components.
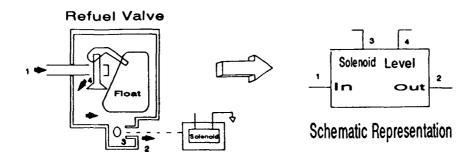
Refuel Valve

Schematic Representation

Figure 2

The first step in model creation is to identify the type and direction of information flow for each component pin/port. Information flow through the component is defined by a set of dependency equations for each of the component's signal input and output pins. For example, consider the level control valve shown in Figure 2. Both the physical and block representations of the valve are shown. The flow of fluid out of the valve (Port 2) requires that fluid be available at the input port (1), that the solenoid be de-energized (3), and that the float not be fully raised (4). The Testability Modeling Language (TML) equations that describe this relationship are shown below:

```
OUTPUT TML EQUATION:
    SIGNAL:        (1)
    CONTROL:       (AND (3 4))
    CONDITION:     (NIL)
    BIAS:          (NIL)
```

All of the information required to create a model is input through a sequence of AutoTEST menus.

Once information flow through the component has been modeled, a set of failure modes for the component must be defined. These failure modes represent the manifestation of physical faults at the component's ports/pins, rather than representing the physical failure. Currently, AutoTEST provides for the representation of three types of faults at each component port/pin. Failure modes may be represented as opens, shorts, and bias-shorts. Essentially, an open fault represents a failure mode which would cause an interruption in information flow at a component input or output. A short represents a failure mode which allows the flow of information between two or more ports/pins along a path which would not be present in normal device operation. A bias-short represents a failure mode which results in the connection of a port/pin directly to power or return. A relative probability of failure (RPF), between zero and one, is specified for each fault defined. The RPF is used in conjunction with the component's mean time between failures (MTBF) to derive the probability of that fault's occurrence.

After the netlist has been read and interpreted, it is translated into a dynamically defined topology knowledge base which contains the components and all interconnection information. The design's components inherit the previously defined model attributes from the statically defined model library. At this point the design is ready for the testability analysis to be performed.

AutoTEST performs testability analysis by identifying how information flows through a design and by determining the effect of the defined

73

faults on the information flow. This information is then used to
derive the FOMs and to apply DFT rules to the design.

AutoTEST identifies four types of information flow through a design:
signal, power, bias, and power-signal. Signal flow may include signal,
control, and conditioning information flow at the component level.
Power flow is associated with a design's input power pins (i.e. current
or voltage sources for electrical inputs and flow or pressure sources
for hydraulic or pneumatic inputs). Power inputs may be used to bias
active components (bias flow) or may flow through components which
control or condition its parameters (power-signal flow). The first
step in the analysis is identification of all possible signal flows
through the design. The method utilized by AutoTEST to make this
identification is detailed in [6].

All the possible signal flows are further divided into flow graphs
(test groups) and signal paths (measurement groups). AutoTEST defines
each valid path between any two adjacent nodes as a directed arc.
Additionally, each arc is assigned a testability weight related to the
component's complexity. All possible signal flows through the design
are described in terms of these arcs. A signal path is defined as a
set of arcs which uniquely identifies a signal flow through a design
which terminates at one output node. So, a set of connector pins and
component states is identified for each signal path. A flow graph is
defined as that set of arcs which results from considering all signal
paths that have the same set of input pins and whose component states
do not contradict one another.

Once all flow graphs and signal paths have been identified, an
accessibility analysis is conducted. This analysis provides a measure
of a design's expected testing complexity. An accessibility analysis
is similar to both the controllability and observability analyses
performed on digital designs. The concept of an accessibility analysis
is based on two earlier works [7,8]. The process utilized by AutoTEST
to perform this analysis is detailed in [6]. The outcome of this
analysis is the calculation of an accessibility weight for each signal
path in the design. These weights are indicative of the degree of
confidence in the information that exists at the associated output
node. These path weights are then used to provide an overall design
accessibility rating (DAR) which quantifies the expected complexity of
testing the design.

Once the operational signal flow through the design has been
determined, AutoTEST provides an estimate of the capability inherent in
the design to detect and isolate all of the defined component failure
modes. This analysis will rely on both the results of the flow graph
analysis and on general knowledge encoded in the models and software.
The development and inclusion of this knowledge in the form of rules
for hydro-mechanical components is one of the major modifications which
must be made to the existing system level version of AutoTEST to create
a version specifically for hydro-mechanical analysis. The testability
rules applied in the existing system and analog version of AutoTEST are
used to identify components or configurations of components known to be
difficult to test, including feedback loops, and to provide test point
and break point location recommendations.

Each component fault is analyzed to determine its effect on signal flow
and the output pins at which it can be detected. This portion of the
analysis considers the immediate impact the fault will have on the flow
graph. For example, if a signal input pin is opened the primary effect
of that fault is to interrupt all flow graphs which include that pin.
The fault will be observable at the associated output pins downstream.
Some faults will affect the operation of components upstream from the
fault. If a fault is on a pin attached to a signal node, AutoTEST will
utilize encoded rules to determine if the fault will affect the
operation of any other device connected to the node. If so the
analysis will continue to look upstream until a node is reached which
will not be affected by the fault. Observation nodes for these
secondary effects are determined by finding all signal paths which
include each of the faulted nodes. A fault table is then created which
associates the faults with the output pins at which they may be
observed for each flow graph. Ambiguity groups, FOMs, and test
strategy are determined using the resulting fault table and algorithms

74

derived from WSTA and incorporated into AutoTEST.

Conclusions: Each of the analysis methods discussed can be used to perform some of the tasks associated with a testability analysis of hydro-mechanical system designs; however, there are also some disadvantages associated with each method. Throughout its development, AutoTEST has incorporated features intended to alleviate the concerns associated with the use of dependency and simulation models for the purpose of testability analysis. A summary of these features is provided in Figure 3. The incorporation of these features also enhance AutoTEST's compatibility with the concurrent engineering approach to design.

TESTABILITY ANALYSIS FEATURES	AutoTEST	Dependency Modeling	Simulation Modeling
Minimal Design Detail Needed To Provide Accurate Results	Yes	No	No
Uses Application Based Rules:			
For Fault Propagation	Yes	No	No
For Design Recommendations	Yes	No	No
Component Libraries Used	Yes	No	Yes
Models Use Standardized Fault Universe	Yes	No	Undeveloped
Provides Figures Of Merit	Yes	Yes	No
Recommendations Easily Applied For Iterative Analyses	Yes	Yes	N/A

Figure 3

The methods utilized within AutoTEST to simplify and standardize the testability analysis process include data parsers (presently EDIF and ICCA), the use of existing and widely accepted WSTA algorithms for FOM and test strategy generation, the explicit representation of faults, and the use of object-oriented characteristics. The use of object-oriented characteristics allows for the development of component libraries which can provide model consistency among users, the development and application of fault propagation rules to specific component classes, and the application of DFT rules based on the inclusion of specific component classes in the design. The explicit representation of fault manifestations, separately from the operational model of the component, allows the user to develop component models with the emphasis on component operation. Component faults need not be implied in the definition of model attributes as is done in dependency models by the creation of aspects. This allows AutoTEST models to provide a more detailed fault representation early in the design process than automatically generated topological dependency models which include a single aspect for each component output.

AutoTEST also provides identification of those components or configurations of components known to be difficult to test, including feedback loops. Recommendations for test point and break point locations based on node accessibilities are provided to improve the design's overall testability. Additionally, AutoTEST provides the capability to perform "What-If" analyses. This capability allows the user to make iterative changes to the design and rerun the analysis. The design changes implemented may be defined by the user or recommended by AutoTEST.

References:

1) Nolan, M. "Can Industry Standards Satisfy DoD and Service ATE Standardization Goals? Can a Standardization Policy be Derived to Satisfy Everyone?", AUTOTESTCON Conference Record, pp. 435-445, 1992.

2) General Dynamics, Fort Worth/McDonnell Aircraft Company, "IDD Phase I- Final Report on Evaluation of Dependency Modeling and Advanced Diagnostics", 1992, Navy Contract No. N00019-88-G-0051.

3) Ofsthun, S. C. "An Approach to Intelligent Integrated Diagnostic Design Tools", AUTOTESTCON Conference Proceedings, pp. 319-329, 1991.

4) Ofsthun, S. C. "Linking Artificial Intelligence and Computer Aided Engineering to Analyze the Testability of Electronic Designs", AIAA Computers in Aerospace, pp. 679-685, 1989.

5) Wilmering, T. J. "AutoTEST: A Second-Generation Expert System Approach to Testability Analysis", Proceedings of ATE & Instrumentation Conference West, pp. 141-152, 1992.

6) Kuhns, F. "Automating Testability Analysis of Analog Circuits and Systems", AUTOTESTCON Conference Record, pp. 225-231, 1992.

7) Longendorfer, B. "Computer-Aided Testability Analysis of Analog Circuitry", AUTOTESTCON Conference Proceedings, pp. 122-125, 1981.

8) Unkle, C. R., Spillman, R. "An Advanced Application of the PCB Testability Design and Rating System", AUTOTESTCON Conference Proceedings, pp. 401-406, 1983.

A SYSTEMATIC APPROACH TO ELIMINATING
FAULTS IN SPACE FLIGHT MECHANICAL HARDWARE

Stephen W. Daudt & Ron King
Ball Aerospace Systems Group
Boulder, CO 80301

Abstract: This paper presents a systematic and unique method for identifying, analyzing and eliminating faults in the early stages of program design. At the beginning of a program, when time and money resources are restricted and test results are limited, effective reliability work can be accomplished. Cost effective reliability growth of developmental design can be made using the principles of concurrent engineering, classical reliability analysis, decision theory by quantitative analysis, deterministic modeling, and limited testing.

Key Words: Concurrent engineering; fault identification; fault elimination; reliability growth; risk factor; mitigation factor

Introduction: Concurrent engineering is one tool within the Total Quality Management (TQM) process that allows for trade-off studies (and hopefully optimized decisions) among several engineering disciplines (e.g.: packaging, thermal, mechanical reliability) that will result in enhanced customer satisfaction.

Over specification and designing with maximum safety margins are not acceptable approaches to enhancing a product with today's limited resources (e.g.. dollars for research, time to production). Reliability is one of these parameters that reflect customer satisfaction (i.e. failures, performance) but cannot be maximized at the expense of design features and increased cost.

In the initial design work on a program we can apply classic reliability tools in a systematic way to enhance customer satisfaction (i.e. reduce fault probability of occurrence) with a minimum investment.

Reliability must be performed as an integral part of the design process (i.e. concurrently engineered). The advantages and the disadvantages of reliability work during early design are shown as follows:

Advantages	Disadvantages
Flexible design	No test data
Less costly to implement	Less design definition
More options possible	Trade-off studies can be costly
Changes easier to incorporate	

The disadvantages show the need for a structured, systematic approach. The methodology presented here consists of a two stage process:

1) Identify and assess probability of major faults

2) Verify assessment and eliminate or reduce faults

A logical series of steps shown in Figure 1 provides a straightforward method of analyzing and improving reliability. We start with a clearly defined system and analyze failure rates to whatever level is initially practical. Next we study faults using a fault tree. Then we compute risk and mitigation factors to see what part of the system can be improved and how. We then analyze and test the system. Finally improvements are made and if the development cycle is long enough the process can be repeated.

Stage 1 | Stage 2

| Define System | Fault Prediction | Fault Tree | Risk Factor | Mitigation Factor | Analyze & Test | Improve System |
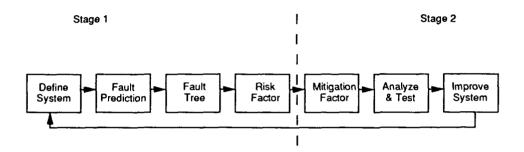
Figure 1   Sequence for Improving Reliability

Each of these steps are discussed in more detail as follows:

Stage 1

Fault Prediction: Reliability work can begin as soon as a system is defined. Fault predictions can be made very early in the program cycle. The probability of occurrence of a fault is intangible (i.e. it cannot be weighed, measured, etc. at a single point in time), so predictions or forecasts are required. Reasonable estimates can be made for equipment that has field experience or related test data, but no heritage state-of the-art development hardware with presents a problem. Obviously there is insufficient time or money to perform comprehensive life testing to obtain this data, so estimates with uncertainty factors are needed. These are best utilized when the unreliability is roughly the same throughout the units (segments, components, functions) of the system. If failure rates are not about the same, the unreliability will reflect the accuracy of the dominant unit. If possible the system should be broken down in to comparable fault rate components. This can be accomplished using numerous techniques including FMEA, fault tree, and math modeling techniques.

Significant Fault Determination: Many failures can occur in a system. Some are catastrophic and others just degrade performance. Not all unit level faults (i.e. failure

78

modes) will result in an identifiable fault at the system level. Many faults simply result in graceful degradation of primary features or loss of secondary features. In initial reliability analysis we must uncover all significant failures.

The fault tree analysis is the best tool for easily finding these major faults. Cost effectiveness is achieved since there is no wasted analysis time on determining secondary effects. Also, customer satisfaction is enhanced since the selected fault types may be seen from an end use viewpoint. Fault tree analysis identifies multiple and dependent faults using a combined bottoms up and tops down analysis. An example of a fault tree is shown in Figure 2. From the discussion above, the fault tree should be developed with an awareness that comparable size failure nodes should be developed. If one node is expected to have a much higher failure rate than others, it should be broken down further.

Initial Fault Probability of Occurrence Determination: A list of significant faults is tabulated based on the fault tree analysis. The bottom level of the tree makes up each line item in the table. This forms the basis of a "first cut" fault probability of occurrence. The confidence in the "first cut" is a function of the element and the existing data base. The further the item varies in design for the item tested, the less confidence there is in the fault probability number. There is also a confidence in the data base (number of samples, mean, variance, etc.). It is desirable to "rank" all the faults to determine how to best allocate resources for improving reliability. A simple quantitative analysis approach may be used as illustrated in Figure 3. A highest risk factor (RF) is identified as having the highest probability of failure and the lowest confidence in the estimate.
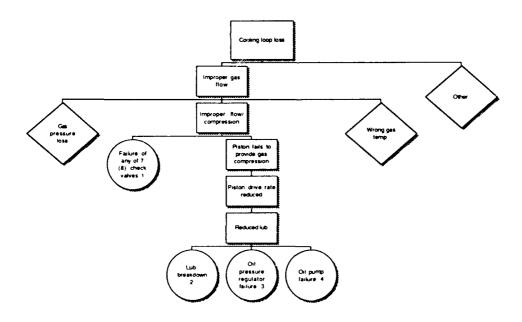

Figure 2   Fault Tree Example For Refrigerator

79

```
Failure        >0.2      5    4    3
  Rate        0.1-0.2    4    3    2
              <0.1       3    2    1

                         L    M    H
                          Confidence
```
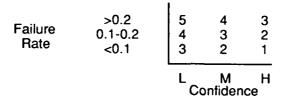
Figure 3   Risk Factor Matrix

All faults with a risk factor of 5 should be given the most attention since they represent the biggest uncertainty. Then those with 4 rating should be considered and so on.

Fault Type Assessment: The first way to examine faults is to assess what failure mechanism might occur to cause the fault. This could be determined by inspection, experience, or analysis by design specialists. Design alternatives then can be considered to eliminate or reduce the failure. Some options include adding redundancy, choosing different materials, or making design enhancements.

To be effective, this must be accomplished early in the design phase (i.e. concurrent engineering). If the failure mechanism cannot be easily identified then small scale testing and analysis may be appropriate. To determine what testing or analysis to apply, the type of failure must be determined. Four major categories are considered.

Infant mortality   -   Usually associated with workmanship, poor design, electronics, etc.

Low cycle fatigue   -   Occurs with few cycles, bending stress

High cycle fatigue   -   Coefficient of elasticity not exceed, more of a shock pulse, long term wear out

Random   -   Not associated with workmanship or wear out

These failure types provide a basis for the next stage which evaluates the various failures.

An example of the fault table that is generated from the first stage of analysis is shown in Figure 4. Having concentrated on identifying faults, we move on to the second stage where the emphasis is on failure reduction.

80

| Event | Qty | Histor. rate | Allocat rate | Risk Factor (RF) | Failure Mode | | | |
					Low Cycle Fatigue	Infant Mortality	High Cycle Fatigue	High Time Wear
Bearing, Ball (other) failure	6	0.06	0.06	1				X
Plumbing	50	2.08	0.1	4			X	
Coupling	4	0.4	0.05	1		X		
Pistons & cylinders	4	0.8	0.1	3	X			
Fail. of any 7 (8) check valves	7	3.5	0.5	5			X	X
Unloader valve fails open/close	1	0.3	0.09	2	X		X	
O-ring	5	0.08	0.05	2				X
Oil pressure regulator failure	2	0.6	0.2	4				X
Oil pump failure	1	3	0.15	4				X
Check valves	2	0.6	0.1	4			X	X
Plugged oil filter	1	0.08	0.08	1				X
Motor failure	1	1.02	0.2	4	X		X	
					Time compress. /accel. testing	Life test	Accel test /modeling	Accel test

Figure 4   Fault Table Example For Refrigerator

Stage 2

Failure Mitigation Plan: The first step toward mitigating failures is to estimate how failures may be reduced or eliminated. A group session of appropriate lead engineers and managers is held to review each failure identified in the fault table. (See example shown in Figure 4.) Ideas to mitigate the failures are discussed. A table, like the one used to rank risk factors, is used to rank failure mitigation. See Figure 5. Here the costs of reducing failures can be weighed against the probability of doing something about the failure. Items with the highest mitigation factor tell where the greatest payoff potential is. Resources (time, money, and manpower) should be spent according according to the mitigation ranking to get the biggest improvement in reliability. Brainstorming ideas are recorded that show how the failure mitigation may be implemented. In our example, Figure 6 shows that hiring a person specializing in valve design can help reduce failures. Many methods can be used to reduce failures. Some listed in the figure involve improved design, production, analysis, and test. Improved design and production efforts are generally lead by production test or quality assurance engineers. Analysis and test programs generally are led by reliability engineering.
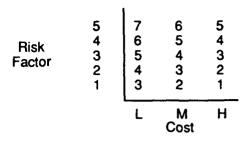
81

		L	M	H
Risk	5	7	6	5
Factor	4	6	5	4
	3	5	4	3
	2	4	3	2
	1	3	2	1
			Cost	

Figure 5   Mitigation Matrix

Event	Allocat rate	Confid. alloc.	RF	MF	Justification & Improvement
Bearing, Ball (other) failure	0.06	H	1	1	Light load, slow speed, wave washer
Plumbing	0.1	L	4	6	Low stress, tests, extra inspection
Coupling	0.05	H	1	3	Use spline
Pistons & cylinders	0.1	H	3	5	Pistons, keyway
Fail. of any 7 (8) check valves	0.5	L	5	6/5	Consultant, extensive design tests
Unloader valve fails open/close	0.09	M	2	4/3	Few cycles, redundant design possible
O-ring	0.05	M	2	4/5	Small temperature variation
Oil pressure regulator failure	0.2	M	4	4/5	Same as check valve use capillary tube
Oil pump failure	0.15	L	4	5	Special alignment in production
Check valves	0.1	L	4	4/5	Oil lube. small cycles
Plugged oil filter	0.08	H	1	3	Oversize
Motor failure	0.2	M	4	5	Rate excludes bearings, kapton insulate

Figure 6   Example Failure Table

Analysis Program: Some failures are mitigated by analysis work. Failures that are impractical to test must be analyzed. Items that are too costly or time consuming to test are good candidates for analysis. Deterministic modeling is used to predict failure modes of these components, determine sensitivities, and assess margins. Failure modes can often be reduced or eliminated even if faults cannot be verified in test.

Test Program: A test program is often a cost effective method of exploring and identifying faults. Items that can be made to fail can frequently be improved. Tests on low cost items or items that fail soon, are good candidates for test. Tests must be constructed to mimic conditions the part will see in final use. In some cases accelerated, compressed schedule, or harsh environment tests may be developed to realisticaiiy provide early insight on reliability. A combination of analysis and test of some parts may be the best solution.

Summary: A systematic approach to eliminating faults can be started up front in a program. The benefits of beginning the effort early are significant. As the program evolves the analysis is fine tuned and reapplied to get a better handle on the systems reliability. The reliability effort must be proactive and ready to support the dynamic needs of the program. The approach just described can help get complex design programs off to a good start.

THE ROOT CAUSE OF ALL FAILURE
OR
WHEN SHOULD WE STOP ASKING WHY

C. Robert Nelms
Failsafe Network, Inc.
P.O. Box 35064
Richmond, VA 23235

Abstract: It is impossible to prevent mechanical failure without addressing its root causes - impossible. The ultimate focus, therefore, of any failure prevention effort should be on root cause. Once the roots are discovered and addressed, failures will not reappear. But this is not "news!" We all know this - at least, we give lip-service to it. But most of us conveniently side-step the real issue, i.e., "What is a root cause?"

Introduction: Those of us familiar with failure analysis and failure prevention know the importance of the question **"WHY."** When reacting to a failure, the failure analyst must continually ask "WHY DID this occur" until the roots of the failure are identified. When proacting to a failure, the risk analyst must go through the same thought process - this time asking "HOW COULD this occur."

In both cases, the analyst usually makes an intuitive judgement as to "when to stop asking WHY (or HOW)." This "judgement call" defines the level to which root causes are established.

If, for example, the failure analyst is a metallurgist by training, he is most likely to "see" metallurgical roots. Structural engineers will "see" stress and strength roots. Managers will "see" organizational roots. In other words, each of us tends to "see" roots at the levels we feel we can influence, or in the areas in which we have knowledge.

Because of this random approach to root cause identification, most of the time we do not ask WHY or HOW to a sufficient level of understanding. We never get to the real roots. Failures reoccur - perhaps not the same, identical physical failure, but certainly something triggered by the same **latent** cause.

The concept of latency is a powerful way of looking at root cause. As the name implies, latent causes lie dormant

within an organization - lurking in the background awaiting a chance to trigger a failure. Latent causes are of two varieties.

One type of latent cause is intensely personal — and it exists within each of us. It addresses the capacity we each possess for choosing "wrong" over "right," fully knowing the difference. For some reason, we sometimes choose wrong - all of us do. This seems part of the human condition - *we* are incapable of *consistently* doing what we know we *ought* to do.

When failures occur on our production lines, or on our aircraft, automobiles, and other things, they often occur because we did not take into account the fact that we're human - and that we all choose wrong over right occasionally.

But latency also addresses another, less personal perspective. In fact, rather than focusing on the person, this type of latency focuses on all those factors *influencing* the person. In essence, this type of latency comes to the rescue of the people intimately involved in a failure. It tends to take people "off the hook."

As humans, we adapt to our environment - whatever that environment might be. We are *conditioned* by the *signals* we receive to adapt to the environment.

Organizations (especially the management within the organizations), send *signals* to their people (often unintentionally). The signals themselves are latent causes of problems, because they alter the perception of right and wrong, good and bad, acceptable and unacceptable. A common example of this is television programming, which sends *signals* across the world which imply acceptable behavior. Organizations and societies are largely responsible for their people's perceptions of right and wrong because of the **signals** they send to their people.

In addition to the signals we send to one another, other latent factors influence the performance of people. The extent to which we are trained for our jobs; the match between our personalities and our job functions; the amount of practice and rehearsal we give ourselves; the respect we have for our leaders; etc. Latency issues are so strong that people often have no choice in the matter at hand - they are forced into certain modes of behavior.

Imagine a world where our questioning processes intentionally and specifically identified the latent causes of our problems.

If we are to tap the gold within latency, we'll have to look closely at two areas: first, our individual

tendencies to *ignore* what we *ought* to do - and secondly, those factors which help *form* our impressions of what we *ought* to do. This paper suggests a specific path on which to travel to help discover these real root causes of our failures:

1. First, we should look at the manifestation of the failure itself, and understand its *immediate causes*. By addressing these causes, we can prevent repeat manifestations.

2. Secondly, we should acknowledge the specific points of *inappropriate human intervention* for each immediate cause, i.e., we must understand how the human intervened by specifying the inappropriate action.

3. Thirdly, we should pinpoint the specific *situation* encountered by the person which led to the inappropriate action.

4. Finally, the investigator must put himself in that situation, then determine *"what, about the way we do business* allowed this failure to occur?"

Look at the manifestation of the failure itself, and understand its immediate causes: Whatever the "failure", it always manifests itself in a way in which it can be characterized. Physical failures are the most easily characterized. Physical failure analysts have developed their own set of jargon to explain all kinds of physical fractures.

But failures are not limited to physical fractures. For example, the product being manufactured can have "quality deviations." The people producing the product can have "emotional problems." Organizations can fail to produce a profit, and "go bankrupt." But whatever the "failure," it always manifests itself to our senses.

It is imperative to macroscopically and microscopically characterize the failure - whatever the failure - to the most detailed degree possible. The experienced investigator knows that the most important "clues" are in the details. The details "talk" to the investigator, explaining to him precisely "what went wrong."

A simple example will be presented to clarify some of these points. Several years ago, an air compressor failed unexpectedly and catastrophically. The main compressor shaft had fractured. After gathering appropriate evidence, the investigator pinpointed the physical cause of the failure.

A bluish discoloration was found on the shaft's main bearing journals. The compressor shaft was also "bent." Both of these clues suggested that the main bearing journals had overheated.

WHY did the bearing journals overheat? One of the possibilities was a lack of lubrication. By sampling the lubrication, the lubricant was found to be contaminated with water. In fact, by sampling the lubricant in its reservoir, it was found that the water level was high enough to have been drawn into the bearings – displacing the intended lubricant.

The above explanation describes the **immediate causes** (or, in this case, the *physical* causes) of the failure.

Acknowledge the specific points of inappropriate personal intervention for each immediate cause, i.e., understand how people intervened: Every failure manifestation is "triggered" by a point of inappropriate personal intervention. This fact is one of the "tests" used to confirm that the immediate causes of the manifestation are adequately understood, i.e., the investigator continues to ask "why" until he finds the specific point(s) of inappropriate personal intervention. The investig 'or is interested in specifics – specific acts of *omissio or commission*.

Note the term used to describe this essential milestone: **point of inappropriate personal intervention.** No mention is made of *"error."* Very often, the person does not make an error – he does exactly what he has been told, yet it *was* inappropriate. Extreme care is taken to pinpoint the *act* – not the person, but the act.

Continuing with the compressor example, the investigator would naturally ask "WHY was water in the lubricant?" In this case, the investigative team hypothesized that either the water had entered the lubricant at the compressor, or water entered the lubricant in storage. The hypotheses themselves drive the search for additional batches of evidence. In this case, both hypotheses were checked, with the evidence pointing to the storage warehouse.

The investigator found that the lubricant was being stored outside. Most importantly, the cap on top of the lubricant barrel was missing, exposing the lubricant to the environment. Considerable water was found within the partially-filled barrel.

Since the *act* of "taking off" and "putting on" the cap is performed by a person, the investigator acknowledged that he had found the *point of inappropriate personal intervention* – someone did not replace the cap.

86

Pinpoint the specific situation encountered by the person which led to the inappropriate action: As stated above, the investigator must first understand precisely what the person did (someone did not replace a cap). Next, the investigator must understand the *situation* which led to the missing cap. Was the person filling the lubrication barrel, simply *forgetting* to replace the cap? Was he siphoning-out existing stock? Was he sampling oil for the quality assurance lab? Or, was the barrel received with a missing cap? In this case, it was found that the cap was left open after an operator had removed a sample for the quality assurance lab.

By seeking and finding the precise *situation* which led to the inappropriate act, the investigator can begin exploring the circumstances and mindsets responsible for the inappropriate act.

Finally, the investigator must place himself within the management system – putting himself in the situation he has pinpointed. He must determine "what, about the way we do business encouraged this inappropriate action?": The placement of the investigator into the "shoes" of the other person is vital – for without this transference, any attempt at understanding the human element is impossible. **The investigator's humanness is the only means of understanding another persons humanness.** A calloused, insensitive investigator will find many reasons to blame people for the failure. But the humane investigator will understand what happened to such a degree that he will be convinced he'd have done the same thing under similar situations.

Following the compressor example a bit further, the investigator found that the quality assurance operator was expected to take 16 samples per day from various locations throughout the plant. This consumed the operator's time. In fact, it was physically *impossible* for the operator to properly sample this many fluids.

The unintended **signal** being broadcast by management was: "Don't worry about doing things right – just make sure you do everything on your lists. If you have to do things half-way, that's okay." This signal was a latent cause of the compressor failure.

In response to this signal from the organization, the operator judged which of his samples were most important, and spent most of his time on these select few sample points. Since the barrels of lubricant to be sampled were placed outside, and had been there for years, it appeared as if they were not considered very important.

87

In addition, the intended cap for the sampling port had been missing for months. In its place, someone had previously pushed a wad of paper in the port to plug it. This was not an unusual practice, as 10 of the 50 barrels of lubricant were also plugged in the same manner. Again, the signal being received by the operator was: "Don't worry about doing things right, just do everything halfway."

Therefore, the operator sampled the lubricant, then replaced the wad of paper. But he pushed it too hard, and it went all the way through the port and into the barrel. The operator said to himself, "Don't worry about doing things right, just do them halfway," and did not bother to take the time to re-seal the port with another wad of paper or the proper cap.

Now that the investigator understands the predicament in which the person finds himself, it is helpful to use the **Self vs. Conscience Model** (see Figure 1) to help solidify a root cause understanding.

The Self versus the Conscience: It can be helpful to view a person as if he came in two pieces - self and conscience. The self perceives **"situations."** It **filters** and transforms the situations into **desires**, then develops a **plan** to fulfil the desires, then translates the plan into **actions** - mostly in the form of body motor functions.

Separate from the self is the conscience. The conscience observes and evaluates the output of the self. **First, it evaluates the desire** produced by the self by suggesting whether or not the desire is worthy (Is it a worthy "end?"). Secondly, the conscience **evaluates the plan** produced by the self - again by suggesting whether or not the plan is worthy (Do the "means" justify the "ends?"). But the conscience has no direct control over DECISION-MAKING. It only acts as an independent, outside "advisor." **The final ability to decide resides within the self.**

When something goes wrong (a failure), the self's output is flawed (inappropriate personal intervention). To be more specific, either the desire itself, the plan to fulfil the desire, or the ability of the body to actuate the plan is inappropriate. This is restated below for clarity:

The Desire Might Be Inappropriate

The Plan Might Be Inappropriate

The Ability of the Body to Accomplish the Plan Might be Inappropriate

88

The sampling operator arrived at the lubrication barrel, removed the wad of paper from the sampling port, and drew a sample. As the operator attempted to re-seal the port with the same wad of paper, the paper fell into the barrel.

Picture yourself in the shoes of the operator. He probably thought, "oh no!" This "oh no" feeling is one of many different kinds of situations our selves must deal with. The sampling operator perceived this "situation." His self's "filter" attempted to force a contrast between the situation "as-is", and the situation "as-desired." But this particular operator's self saw no difference between as-is and as-desired. His self wanted to be able to finish the required number of daily samples, and did not want to bother with extraneous chores.

The resulting desire (or lack of) was inappropriate, resulting in the compressor failure.

But the operator cannot be blamed for not having the desire to plug the sample port! The operator's **filter** did not allow the operator to perceive this desire. If another operator had been in that same situation, his self *might* have **filtered** the situation differently. But then again, he might not have - it depends entirely on the operators **filter**.

This filter requires additional discussion. The self develops slowly as it encounters varying life experiences. Our filters are formed as a result of these life experiences - by the signals experienced by the self. **Filters are an indistinguishable part of the self.**

As we learn about the causes of our failures, and see that we can trace them to our filters, we also begin to understand that **an organization might be able to influence this filter** - by managing the signals experienced by the selves. In other words, we (as humans) have control over the many of the signals which form our filters.

But when organizations or people neglect or ignore the signals they send to one another, our filters cannot help but degrade. As humans, our tendency is to send out **"expediency signals."** These unintentional signals tell everyone around us to "Do it the quickest way you can, and look out for yourself because no-one else will." Of course, we know better than this - we know this is not the best way to think or act. But we continually have to remind ourselves of this, or the self and its expedient preferences take over.

Expediency signals are insidious! For example, if it is perceived "okay" to do things half-way, everyone's filter will gradually change until no-one will desire to do things

right. Although this is never an intentional signal, it is never-the-less the prevalent signal if left to chance.

If the operator's filter would have created a difference between the as-is and as-desired situation, the operator would have acted. But in the above example, the operator did *not* act. He did not *desire* to act. His filter did not create a desire. **The filter caused the failure - not the operator, but the filter.**

But the operator is not "off-the hook" yet. We have yet to consider the operator's conscience. His conscience always has a "say." His conscience told him that he *ought* to replace the paper with either another piece of paper, or a real cap.

Many experts say that we all have the same conscience - that it's a timeless and cultureless "knowing," similar to being instinctive. These same experts say that as opposed to people having varying consciences, their equal consciences have varying strengths. It is as if our consciences "talk" to us. Some of us hear the voice loudly - others hardly hear it at all. But whatever the strength, the conscience is uniform across all peoples.

Most importantly, however, is our human ability to choose the advise of our conscience - even if it makes absolutely no logical sense - even if it places us in life-threatening danger. **We can either choose to listen, or choose to ignore our conscience.**

In our example, the operator had a choice to make at this point (as we all have this choice to make in all the situations we encounter in life). He could either have listened to his conscience, and replaced the sampling port cap - or he could ignore his conscience. **The operator ignored his conscience.**

The effect of ignoring conscience could be deadly - not literally in the case of this compressor, but perhaps literally in the long run. As the operator chooses to ignore his conscience, with full support of his organization, the little "voice" inside of him gets fainter and fainter. It is as if each decision to ignore the conscience makes the voice fainter. Eventually, the operator will not hear the conscience at all, and will be entirely driven by selfish motives and desires - not caring about quality, hiding information, lying to everyone, and creating the ultimately deadly spiral.

If the reader conscientiously applies the thoughts in this paper, he will likely discover an overwhelming message - **the importance of the conscience in daily business decision-making.** Our personal willingness to listen to our conscience, as well as an organizations insistence that we

90
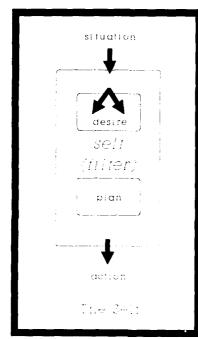
all listen to our consciences, seems paramount to running a
business properly. Even in this trivial example, the
failure of an air compressor was at least partially caused
by an operators refusal to listen to his conscience.

**The writer is of the opinion that all failures are caused
(in the limit) by ths phenomena.** In essence, this appears
to be the root cause of all our failures. Whether it be a
designer, a stress analyst, a materials engineer, a
purchasing manager, a maintenance supervisor, or a janitor
- it seems that all our problems can be traced to someone,
somewhere not listening to his conscience. **If problem
solvers and failure analysts would ask "why", and keep
asking "why" until they exposed this link to our
conscience, we would most certainly be a more productive,
happier, profitable society.**
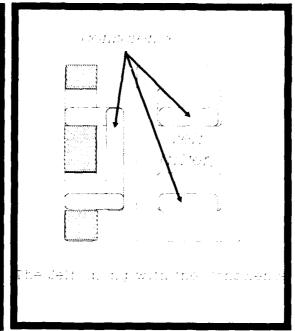
Admittedly, this paper is a broad departure from the
typical discussion of failure prevention strategies. But
attempts to prevent mechanical "failure" which avoid the
"people" issues are missing the point.
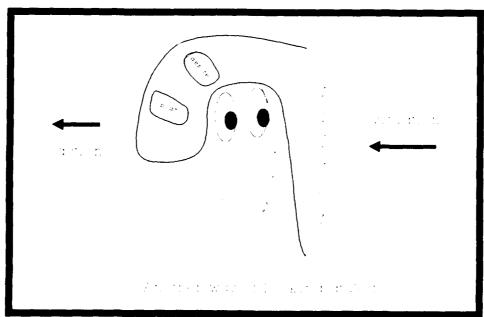
We, our SELVES are the ultimate cause of all failure.

FIGURE 1

NALCOMIS LOCAL AREA NETWORK:
AN INCREMENTAL APPROACH TO JOINING
SHIPBOARD COMPUTER APPLICATIONS

P. T. Smiley, Capt, USN
NALCOMIS Program Manager
Naval Air Systems Command
Naval Air Systems
Command Headquarters
Washington, DC 20362-5160

Joseph Polidoro
Naval Ship Systems
Engineering Station
Philadelphia Naval Base
Building 633
Philadelphia, PA 19112-5083

ABSTRACT: The U.S. Navy is striving to reduce the tedious administrative burden traditionally associated with maintenance and work planning both shipboard and ashore. The transition to computer-based maintenance systems is well underway, with numerous initiatives in progress for molding business process improvement, monitoring equipment performance, diagnosing equipment problems, generating repair recommendations, tracking maintenance costs, inventorying parts, and planning maintenance availabilities. Electronic technical manuals are also being developed. These initiatives cross many organizational boundaries and require cooperation and collaboration between many people and a multitude of diverse organizations. To be effective, the various computer enhancements developed for use in our ships must communicate with each other and with established programs and be compatible with the supporting infrastructure ashore. Our ships need a computer data highway capable of handling the myriad of computer applications, data transmission and electronic communication protocols in place now, currently planned, and yet to be envisioned. This paper will discuss the Naval Aviation Logistics Command Management Information System (<u>NALCOMIS</u>) Local Area Network (LAN), which has been developed and tested by the U.S. Navy to serve the immediate needs of NALCOMIS and be compatible with existing shipboard computers and the increasing computerization of ships in the future. A historical summary of the development of this fiber optic bus is provided, along with descriptions of key technical aspects of the system, and the methodology for fleet-wide implementation.

KEYWORDS: Ethernet; FDDI; Fiber Optics; LAN; Token Ring

INTRODUCTION: The U.S. Navy is getting computerized. Massive amounts of documents that currently laden our ships are being reissued on CD-ROM. Maintenance functions are being streamlined as computer generated work packages replace hand written work requests and maintenance tracking reports. Additionally, performance data logging and trending is being accomplished by automated diagnostic systems, and the expertise of equipment specialists is being captured by expert systems that identify common faults and make repair recommendations. The applications already established or under development are many and varied. However, they all have two things in common: they must reside on a computer and, to be truly effective, they must interact with each other. The Navy cannot afford to procure and maintain separate computer assets for each application.

APPLICATIONS: Several computer based applications are developing. The Machinery Condition Assessment System (M-CAS) uses real-time data acquisition to monitor systems such as boilers, main engines, and electrical generators. These systems will be tied to automated diagnostics and expert systems to provide trends, alarms, machinery history data, repair recommendations, maintenance, and management functions. Expert systems are computer programs designed to be a "Technician in a Box." Expert applications use deductive reasoning based on a database of knowledge, procedures, and feedback to aid a technician in completing a task. Expert systems for these applications are comparable to an interactive maintenance guide. To supplement M-CAS and expert system applications other information usually stored on paper will be digitized, distributed on CD-ROM, and managed on the computer system. This includes tools and parts information, technical manuals, drawings and blueprints.

These applications will build upon existing shipboard applications such as The Shipboard Non-Tactical Automated Data Processing Program (SNAP II) which generates work requests, builds work packages, tracks supply parts, and assists in other administrative functions.

LOCAL AREA NETWORKS: Many of these applications are computationally intensive and resource hungry. If it is desired that these applications run on various work stations at different locations it makes sense to implement the applications over a LAN. A LAN allows for the reduction of computer-power redundancy. Instead of every workstation having the resources needed to run the desired applications, the required computer-power can be networked from centralized servers. The workstations handle the interaction. The LAN allows for the sharing of resources such as applications, data, printers, modems, and storage. It handles E-MAIL, automated back-ups, and security monitoring. It is a key tool in the task of managing, maintaining and updating software and workstation configurations.

THE NALCOMIS LAN: To support these functions a capable, efficient, and upgradable LAN is needed. The NALCOMIS LAN specifications were developed with these requirements in mind. The NALCOMIS fiber optic LAN for shipboard use is being designed by the Naval Undersea Warfare Center, Norfolk, Va and the Naval Surface Warfare Center, Carderock Division, Naval Ship Systems Engineering Station, Philadelphia, Pa (NAVSSES). The system meets current federal standards for Local Area Networks and is designed to support expected shipboard computer applications as the Navy enters the 21st Century.

NALCOMIS is based on a fiber backbone compatible with the Fiber Distributed Data Interface (FDDI) being developed in Accredited Standards Committee (ASC) X3T9 which is chartered to develop computer input/output (I/O) interface standards. For a large network FDDI provides for a performance factor an order of magnitude higher than a typical Ethernet LAN. FDDI inherently provides for maximum upgradability, mixing of manufacturers' equipment, live maintenance, and high survivability.

94

A key facet of the LAN is its use of fiber optics. The high bandwidth of optical fibers allows the use of a bit-serial transmission protocol that significantly reduces the size, cost and complexity of the hardware required by a network. In input/output channel applications, for example, single duplex optical connectors can realize the data throughput of eight 48-pin coaxial cable connectors. Applied to the LAN as a whole, fiber optics greatly enhance the capabilities and capacity of the system.

The NALCOMIS LAN's unprecedented reliability can be largely attributed to its ring configuration. Use of a ring offers several advantages. Reliability and survivability of the LAN are greatly enhanced and hardware installations are simplified. The ring readily accommodates the use of optical fiber and offers some significant advantages in the ease of initial configuration and reconfiguration as the needs of a network change. A ring inherently imposes no restrictive logical limit on the length of ring links, the number of stations and the total extent of the network that can be accommodated.

THE NALCOMIS DESIGN: The excellent characteristics of NALCOMIS are attributable to its underlying architecture. NALCOMIS combines the optimum network configuration and protocol with high speed fiber-optic communication links. To understand the advantages of the NALCOMIS LAN one must understand the underlying limitations of typical LANs in use today. The most commonly used today are copper-wire Ethernet LANs. Workstations on Ethernet communicate on a common wire (see Figure 1) using coaxial cable (COAX) or an unshielded twisted pair (UTP). Typically when one
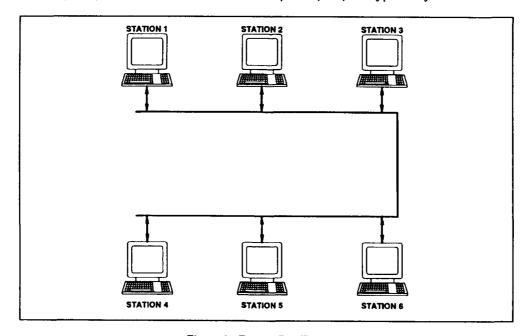
Figure 1. Typical Bus Topology

station wants to send a message to another station it waits until the line is clear then transmits a packet of data. This packet of data contains destination station address, error checking bits, data, etc. All stations monitor the transmission but only the destination station uses and acknowledges the data.

The major disadvantages to this architecture are:

- As more stations are added to the common bus the performance of the network decreases rapidly. This necessitates the breakup of the bus into smaller busses connected by routers. This adds cost and complexity. Compounding this, network utilizations only approach 30 percent of the theoretical maximum load capacity.

- The total length of any one bus is limited to approximately 1000 feet.

- There is no bus arbitration. If two or more stations accidentally broadcast over each other the transmissions are resent some short and random time interval in the future. While the lack of bus arbitration greatly reduces protocol overhead there is no guarantee that any particular station will get a chance to broadcast in a timely fashion.

- A failing station can tie up and confuse a network causing the whole system to crash.

The NALCOMIS LAN architecture addresses these problems and adds significant features which are detailed in the SPECIFICS section of this paper. NALCOMIS uses a ring architecture (see Figure 2) with point-to-point fiber optic data links. A special bit-pattern called a "token" is transmitted from station to station, circulating around the ring. When one station wants to send a message it attaches its message and destination address to the circulating token. Multiple tokens can be circulated thus taking advantage of the pipelined architecture of the ring.

This architecture leads to the following advantages:

- Network utilizations exceeding 90 percent are readily achievable. Rings are insensitive to load distribution and the performance is not degraded significantly by the presence of inactive stations.

- There is no logical limit to the number of stations or the total length of the ring.

- Rings provide time-bounded access delay for data transmission under all conditions.

- Failing stations can be isolated remotely.

If a ring can cost-effectively yield great performance increases and offer other advantages as well why aren't rings in more common use? The need for such connectivity and peak data rates was not apparent in the marketplace until recently. It is only after the demand was created that
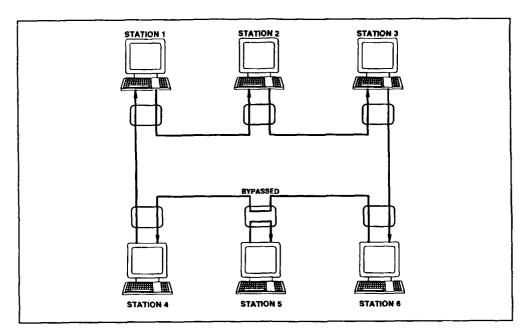
Figure 2. Typical Ring Topology with Bypass Capability

standards were developed and hardware was designed. The price of this hardware is falling rapidly as token ring LANs become more popular.

Specifics: As mentioned before, the NALCOMIS LAN specifications were built on FDDI. The FDDI specifications incorporate the architecture, transmission medium, and protocol. FDDI utilizes a token ring architecture and employs optical fiber as a transmission medium. Each ingredient in FDDI compliments each other resulting in many advantages over other networks. This section details each of these ingredients.

Fiber Optics: Because of the nature of fiber optics, and the fact that each connection in a ring network is a dedicated link, the transmitters and receivers can use an extremely wide bandwidth. This allows very high baud rates due to a better signal to noise (S/N) ratio. Because of the large bandwidth, data can be transmitted serially. Serialized data communications simplifies hardware. The transceiver hardware need not be duplicated as is required with parallel connections. The transceiver need not manipulate signal amplitude, phase, and frequency in an effort to squeeze more data through a band-limited channel. Dedicated links do not provide channel sharing capabilities as are used in other systems, therefore the hardware need not multiplex data via time and frequency slicing.

In general fiber optics has these advantages over copper wire:

Cost
Optical fiber is the least expensive wiring option for many network applications when factors such as life of the network and cost of upgrading are included.

97

Strength

For the same diameter, glass fiber is stronger than steel. With an average tensile breaking strength of 600,000 lbs./inch^2, fiber optic cable exceeds the strength requirements of all of today's communications applications. Fiber is designed to have a long life expectancy. Glass fiber is an extremely stable material.

Increased Data Throughput

For data rates greater than 100 megabits-per-second (Mbps), fiber optic cable is the only medium that can be used reliably. Fiber optics offers a higher data bandwidth and transmission distance than copper wire. The data transmission capabilities of copper wire are well defined and the development process is mature. Gains in technology and improvements in the manufacturing process will improve the characteristics of fiber optics transmission lines in the years to come.

Immunity to Electro-magnetic interference/transmission

Fiber optic transmission lines are neither effected by electro-magnetic interference nor do they emit EM energy in the radio frequency spectrum. Fiber is immune to lightning strikes and the resultant surges that can damage connected equipment.

Weight/Size

The weight and size of transmitters, receivers, and transmission lines for fiber optics is much better since fiber optics can transmit large amounts of data in a serialized fashion.

FDDI: Industry leaders of the American National Standards Institute (ANSI) subcommittee X3T9 developed a LAN in an effort to standardize high-speed optical fiber LANs. The FDDI standard is dedicated to the comprehensive implementation of communications through fiber optics. FDDI at 100 Mbps is faster than Ethernet (IEEE 802.5) at 4 or 16 Mbps, and delivers higher performance and reliability. It is slated to support an 80 percent sustained bandwidth compared to 30 percent on Ethernet copper wire LANs.

Initially, FDDI was a single 100-Mbps Token Ring. The main problem with this design is that if a break occurs in the ring, the entire system goes down. To reduce downtime, the FDDI standards committee developed a dual ring with built-in redundancy. Based on the dual counter-rotating Token Rings (see Figure 3), FDDI networks can bypass hardware failures. Any failure on FDDI dual rings can be isolated, keeping the remainder of the rings completely active. When the failure is "corrected," the FDDI ring reconfigures automatically. Typically, the primary ring carries data, and the secondary ring is used for automatic bypass and recovery.

To communicate data, a special bit pattern, called a token, is continuously circulated by the FDDI ring. Stations transmit data by capturing the token and sending it on a complete circuit of the network. This is a "deterministic" method because each station is guaranteed token service within a specified time limit.
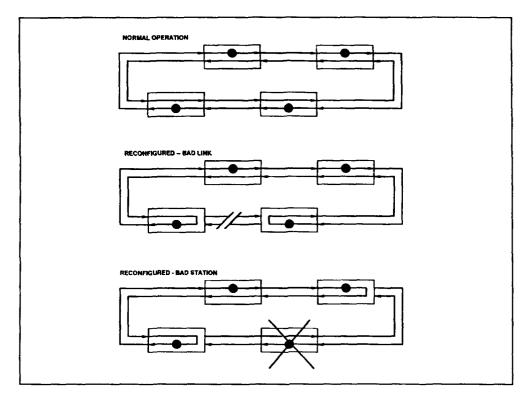
Figure 3. FDDI Counter Rotating Ring Fault Tolerant Examples

FDDI rings support two types of stations; dual attached stations (DAS), which attach directly to the ring, and single attach stations (SAS), such as PCs and workstations. Each DAS has four fiber connections, two to receive and transmit to the primary ring, and two to a secondary ring. A typical DAS can be a concentrator, bridge, router, server, or minicomputer and mainframe. Multiple DASs are linked together to form the network backbone. SAS can be immediately isolated in case of failure without disrupting traffic on the ring.

As a peer-level distributed network, all the DASs in an FDDI ring participate in fault recovery, management, capability, and network initialization. Internal DAS timers and logic control resolution of all ring failures provide bypass handling [1].

Key advantages of FDDI are:

Reliability, survivability, and maintenance
A ring will still operate even when individual stations or a portion of the network is not functioning. Part of the network can be taken down at will without interrupting the rest of the network. In other networks, such as Ethernet, a failed workstation can bring down an entire network.

99

Easier configuration and reconfiguration
Failing stations can easily be isolated. Stations can be added and deleted without adverse impact to existing ring traffic.

Deterministic
A station is guaranteed to have the opportunity to transmit data within the time it takes for a token to circumnavigate the ring.

Simplification
The point-to-point connections allow for a simplified hardware and protocol design. It allows easy mixing of different manufactures' equipment. Different transmission media (such as copper and wireless) can be used on different legs of the ring. Areas of the ring with the most data traffic can be easily upgraded to handle more data without upgrading the entire network.

Circuit and packet switching capabilities
Point-to-point connections bring with it the ability for true circuit and packet switching capabilities.

IMPLEMENTATION: Implementation of the NALCOMIS LAN is a three step process. A proof of concept prototype, such as USS NASSAU (LHA-4), must be developed and installed. Then the prototype must be tested and evaluated by NAVSSES. Once the test and evaluation is completed, lessons learned will be incorporated, and components installed Navy-wide.

NALCOMIS has extensive fiber optics experience on shore installations (see Table 1). The NALCOMIS LAN is currently installed on several air capable ships using copper wire transmission lines. The fiber optic NALCOMIS LAN will be installed on-board NASSAU during fiscal year 1993. Like the previous installations, the NASSAU LAN will use commercial-off-the shelf (COTS) equipment which meets the environmental requirements for shipboard use. The NASSAU LAN will be integrated with the ship's SNAP I system and will represent the NALCOMIS LAN's first afloat application of fiber optic cables. The NASSAU LAN will be installed by NAVSSES Code 103B under SHIPALT LHA1 708K using the Machinery Alteration (MACHALT) process which is described below. Once tested on the NASSAU, efforts will be made to test other shipboard computer applications (expert systems, electronic tech manuals, etc.) on the LAN.

Since the NASSAU LAN is a proof of concept prototype, Measures of Effectiveness (MOE) will be developed in order to objectively monitor and trend its overall return on investment (ROI); both in real dollars and in material readiness.

T&E/Implementation: A test and evaluation (T&E) plan will be developed by NAVSSES in accordance with Navy requirements [2]. It will evaluate the accuracy and reliability of the prototypes, determine if the system is user friendly, be used to maintain ongoing NAVSSES/ship interface, and be used by shipboard personnel as a training tool. After a 6 to 12 month evaluation period, a final report will be issued for each prototype. These reports will contain cost benefit analyses, risk management assessments and recommendations concerning the applicability and

100

ASHORE	
NAS Norfolk	Under Runway Segment
USMC	Deployable LAN
NAS Miramar	Tuttle-Anselmo LAN
NAS Cherry Point	FO connection design in progress
AFLOAT	
CV-64	FO LAN design completed
CVN-73	Support to FO LAN, GWIS
LHA-4	NALCOMIS-NASSAU LAN in progress
LHD-5	Integrated FO backbone design in progress

Table 1. NALCOMIS Team Fiber Optic Experience

effectiveness of the tools and techniques applied. With these reports in hand, along with feedback received from interviews with shipboard and shore based personnel, NAVSSES will be able to upgrade the original design as required and to proceed to Navy-wide implementation using the Machinery Alteration process developed and executed by NAVSSES. Additionally, implementation of the NALCOMIS LAN will be dovetailed with other initiatives to ensure that the LAN will be compatible with all new computer applications.

MACHALTS: Machinery Alterations are used by the U.S. Navy to effect changes to equipment and systems where the changes are contained within the boundaries of the individual equipment or system and have limited impact on other (external) equipment or systems. A MACHALT is defined as a planned change, modification or alteration to any equipment in service (shipboard or shore based) when it has been determined that the alteration or modification can be accomplished without changing an interface external to the equipment or system; is a modification made within the equipment boundary or is a direct replacement of the original equipment design; can be accomplished without the ship being in an industrial activity; and will be accomplished individually and not conjunctive with a SHIPALT or other MACHALT [3]. The MACHALT Program employs a kit installation concept (Figure 4) that enables equipment changes to be accomplished in an expeditious manner and eliminates them from the formal Ship Alteration (SHIPALT) process. The program has been so effective that NAVSSES now uses the MACHALT process to manage SHIPALTS including LHA-1 708K.
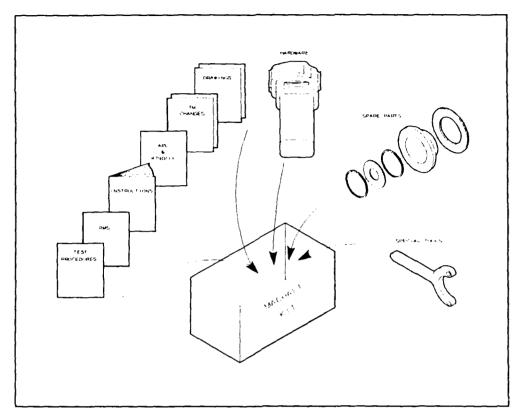
Figure 4. MACHALT Kit Concept

CONCLUSIONS: Continued computerization of shipboard tasks will place an ever increasing demand on the supporting local area network. The NALCOMIS LAN is designed to meet these requirements well into the 21st century. The NALCOMIS LAN is the first step in providing the framework for the implementation of more efficient and cost effective shipboard maintenance programs. This initiative is the first systems command developed installation of a fiber optic network in support of automated information systems.

REFERENCES:

1. Telecommunications, April 1992. "Methodologies for FDDI Test and Measurement," Carol Baker

2. Reference Number S9095-AC-HBK010 / Test Reference Handbook

3. Commander, Naval Sea System Command, NAVSEAINST 4720, "Policy and Procedures for Implementation of Machinery Alterations (MACHALTS) on HM&E Equipment and Systems," 23 March 1983.

POSTER SESSION

Coordinator: Christopher Nemarich
Naval Surface Warfare Center

METALLURGICAL ANALYSIS OF FAILED BOLTS FROM THE M60 TANK RECOIL MECHANISM

Victor K. Champagne
Paul Buckley
Milton Levy
Robert M. Huie
Jonathan B. Fazli
Wayne M. Bethoney
U. S. Army Research Laboratory
Watertown, MA 02172-0001

Abstract: A comprehensive analysis of two failed M60 bolts was carried out at the U.S. Army Materials Technology Laboratory. The bolts broke during installation and were examined for the cause of failure. A total of 69 additional bolts from both inventory and the field were also characterized and tested for comparison.

It was concluded that the bolts examined were fabricated from AISI 8740 steel as determined by chemical analysis. Metallographic examination revealed the microstructure to consist of tempered martensite. More than 50 percent of the bolts contained a sharper than specified head/shank radius. Only one of the additional 69 bolts tested failed magnetic particle inspection due to a transverse crack at the head/shank radius. The torque tests and stress durability tests indicated no bolt failures. Optical and electron microscopy of failed bolts showed topographies and black oxide consistent with the characteristics of quench cracks. The failure mode was attributed to pre-existing quench cracks which should have been detected by the 100% magnetic particle inspection conducted during manufacturing. These cracks propagated during installation causing the bolt heads to sever. Recommendations were provided to minimize future mishaps and to prevent failures in the field. These included improved in-process inspection; the use of dull cadmium plate to mitigate the potential for delayed failures due to hydrogen embrittlement or stress corrosion cracking; an alternate to electro-deposited cadmium plate such as vacuum cadmium plate or ion-plated aluminum; and finally replacing all

existing bolts in the field and in inventory with new or reinspected bolts.

Introduction: The intent of this metallurgical investigation was to isolate the probable failure mechanism of the bolts. A course of action was then formulated and implemented which would prevent defective bolts from entering the inventory and replace those in fielded systems. Several possible mechanisms that caused the bolts to fail were proposed: quench cracking; hydrogen embrittlement; stress-corrosion cracking and overload.

In addition to the two failed bolts (referred to as bolts A and B), 57 bolts were obtained from inventory and 12 were taken from fielded tanks at Fort Knox to be examined conjunctively for comparative purposes. The 69 additional bolts were subjected to magnetic particle inspection for evidence of cracks. Subsequently, the following analyses and metallurgical tests were performed on a number of bolts chosen randomly: chemical composition of the alloy; measurement of the radius at the bolt head/shank interface for indication of excessive stress concentration; mechanical properties and hardness measurements; metallographic analysis for microstructural characterization of the alloy; examination of the cadmium plating for thickness and uniformity; torque testing for maximum torque-to-failure; stress durability testing for externally threaded fasteners which may be subject to any type of embrittlement (such as hydrogen embrittlement induced by cadmium electroplating); scanning electron microscopic examination of fracture surfaces; and elevated temperature exposure tests at the tempering temperature (~1200°F) and stress relief temperature (~375°F) to determine if the black oxide observed on the fracture surfaces of the two failed bolts may have been attributed to a prior heat treatment during manufacturing.

Identification of the Bolt Alloy: The engineering drawing and specifications of the bolt shown in Figure I allow the fastener to be fabricated from any of the following steels: 4140, 4340, 6150, or 8740. Atomic absorption and inductively coupled argon plasma

106

emission spectroscopy were used to determine the chemical composition of the alloy. Carbon and sulfur were determined by the 'LECO combustion method. It was determined that the bolts were fabricated from AISI 8740 steel after comparing the nominal composition of the four alloys with the chemical analysis of the two failed bolts. This low alloy steel is quite similar in properties to type 4130. In the quenched and tempered conditions, the alloy should have a good combination of strength, toughness and fatigue resistance.

Radius Measurement at Shoulder/Shank Interface: Figure I requires the radius to be 0.057, to +0.0000-0.0010. Approximately 50 percent of the bolts did not meet the specification requirement. The radius of these bolts were slightly sharper than specified ranging from 0.058 to 0.060. The increased sharpness could provide sites for crack initiation due to higher stress concentration.

Microstructure of the Bolt: Figure 2 contains a representative micrograph of the failed bolts as well as those from the field and inventory showing the microstructure to consist of tempered martensite, typical of a quenched and tempered low alloy steel. The material was clean with no major inclusions present.

Cadmium Plating Thickness and Uniformity: Metallographic cross-sections of the failed bolts were taken in the shoulder area to examine the cadmium plating. The average thickness of the cadmium plating of failed bolt A was 0.00047 in. and the plate was quite uniform. The plating on failed bolt B was much thinner 0.00016 in. and less uniform. The cadmium plating on both specimens displayed good adherence. In addition several bolts from inventory were sectioned and metallographically examined. These bolts exhibited a uniform cadmium plating thickness of 0.0039 in. Based on thickness measurements, bolt A and those from inventory conformed to the requirements of a class 2 cadmium plating (Federal Specification QQ-P-416 E), while bolt B was a class 3.

Magnetic Particle Inspection: All 69 additional bolts were subjected to magnetic particle inspection for cracks and discontinuities in accordance with MIL-I-6868. Only one of the bolts obtained from inventory failed due to the presence of a crack. Figures 3 shows evidence of cracking revealed by the test using black light photography. This bolt contained a transverse crack near the head-to-shank radius which extended over two-thirds of

107

the circumference of the shank. It was assumed that bolts were 100 percent inspected in accordance with MIL-B-8831B, as required. Since this bolt failed inspection, it can be deduced that either a 100 percent inspection had not been carried out or the crack went undetected during the inspection.

Mechanical Properties: A standard ASTM tensile test was performed on six bolts which had experienced extensive field use at Ft. Knox in addition to the bolt which failed magnetic particle testing. One tensile specimen (0.113 in. in diameter) was fabricated from each bolt head and another from the thread area. The range of values were as follows: Ultimate tensile strength 210 to 218 ksi; 0.1% yield strength 182.5 to 200 ksi; 0.2% yield stength 189.2 to 202.5 ksi: %reduction in area 52.0 to 56.0; and %elongation, 10.6 to 13.2. In addition, tensile tests were performed on 11 inventory bolts listed. These tests were conducted by using actual bolts, and not threaded tensile specimens. All the bolts failed within the threads except for one which failed at the bolt head. The ultimate tensile strength of these 11 bolts were in the range of 198.8 to 212.8 ksi. It should be noted that only 1 bolt of the 11 tested were below 209 ksi. The maximum load-to-failure was between 42,750 and 45,750 lbs. exceeding the minimum specified ultimate tensile load of 39,100 pounds.

Hardness Measurements: A Knoop hardness survey was conducted on the two failed bolts, six fielded bolts from Ft. Knox and six new bolts taken from inventory. The Knoop values obtained were converted to HRC. According to MIL-B-883113, the bolts shall have an HRC of 39 to 43. There was no evidence of a significant hardness gradient between the three groups of bolts tested. However, the two failed bolts and those with a firing history taken from the field exhibited slightly higher hardness values (44.0 HRC to 45.1 HRC) than the bolts from inventory (39.8 HRC to 42.5 HRC).

Torque Tests: Torque tests were carried out on eight bolts from the field and 8 from the inventory employing a calibrated wrench fitted with a heavy-duty socket. The torque-to-failure was within 275 to 450 ft-lb. These values exceeded the minimum torque requirement of 120 to 140 ft-lb. Seven of the bolts failed at the beginning of the threaded section while the remaining bolts failed in the center of the threaded region. Since none of the bolts failed at the head/shank interface, the sharper than specified radius at

108

this interface did not by itself initiate the failure. Generally, the bolts from the inventory exhibited higher torque failure values (Avg. 403 ft-lb.) when compared to the bolts from the field (Avg. 346 ft lb.).

Stress Durability Test: To investigate the possibility that hydrogen may have been introduced into the bolt during the electrolytic cadmium plating operation and may not have been adequately removed by the low temperature embrittlement relief treatment (causing hydrogen embrittlement), a stress durability test was carried out in accordance with MIL-STD-1312-5A, Test 5. A plate fixture was fabricated from 4140 steel and heat treated to HRC 45. This plate was drilled and tapped for 16 bolts. Sixteen load cells were also fabricated from 4140 steel, HRC 45. The load cells were strain gaged and calibrated for load versus strain. Eleven new bolts from inventory and six used fielded bolts were preloaded to 80 percent of the UTS and subjected to a 200-hour test. None of the bolts fractured, and transverse cracks were not observed during inspection of the bolts after testing. Note that MIL-B-8831B specifies that the preloaded bolt shall be maintained at load for only 23 hours without failure. In order to carry out a better statistical sampling of the bolt inventory, 12 additional new bolts were tested, but the duration of test was extended from 200 hours to 400 hours. There were no failures after stressing at 80 percent of the UTS for 400 hours.

Scanning Electron Microscopy of the Fracture Surfaces: The fracture surfaces of the two failed bolts were examined utilizing the scanning electron microscope (SEM). Figure 4 contains a macrograph of the fracture surface of bolt B obtained by light optical microscopy. There were four distinct fracture zones which are depicted schematically in Figure 4. A black crescent shaped area designated Zone 1 was determined to have been the Site of crack initiation. Adjacent to Zone 1 was a grey area, Zone 2, which contained river markings indicative of crack growth. Another light grey area, Zone 3, similar in appearance to Zone 2 was observed and represented faster crack growth as evidenced by a very fibrous mode of fracture. The last crack region, Zone 4, was a shear lip indicative of final fast fracture. Scanning electron microscopic examination of Zone 1 revealed an intergranular fracture surface below the black layer as shown in Figure 5. EDS of this surface showed those elements associated with the steel, as well as oxygen (Figure 6). The black material was later concluded to have been a

high temperature oxide and not contaminants or simple atmospheric corrosion. This type of oxide was similar in appearance to a heat treat scale. Zone 2 was characterized by a mixed intergranular and ductile dimpled topology (Figure 7). Zone 3 also contained this mixed mode of fracture in a very fibrous manner but there was more ductile dimpling than in Zone 2 . Zone 4 displayed a typical shear/ fast fracture morphology as evidenced by shear dimples (Figure 8).

The bolt which had failed the magnetic particle inspection because of a transverse crack at the head/shank radius was subjected to tensile testing to open the crack and expose the two resultant fracture surfaces for examination. The entire surface of this fracture was covered with the same black oxide as bolts A and B except for a very narrow shear lip region and appeared to have the same intergranular fracture mode. There was no evidence of a ductile dimple topology as a result of simple tensile overload. The low load to failure of the bolt (about 1000 lb) obtained during tensile testing, in conjunction with the black oxide covering about 90 percent of the surface and the absence of ductile dimple rupture, indicated the crack area of the bolt encompassed by the black oxide existed prior to tensile testing. This suggests that the cracks in both failed bolts A and B which were also covered with black oxide, were preexisting flaws and not due to the service environment. It appeared, therefore, that hydrogen embrittlement may be ruled out as a failure mode. In order to determine when the crack occurred during the bolt fabrication process, and when the black oxide film formed on the crack surface, elevated temperature tests were carried out. Disc specimens were cut from both the failed bolts, polished through 600 grit SiC paper, and cleaned. One specimen was placed in an oven preheated to 375°F, typical of a low temperature stress relief treatment, and another exposed to 1200°F the tempering range for this material. The specimen heated to 375°F did not oxidize after exposure for 1 hour. However, after exposure at 1200°F for 1 hour, the specimen was covered with a black oxide. The black oxide was present after only 5 minutes of exposure at this temperature. Considering that the bolts were heat treated at 1600°F, quenched, tempered at 1250°F, stress relieved at 375°F, cadmium-plated, and baked at 375°F (+ or -25°F) for 3 hours to prevent hydrogen embrittlement, the cracks most likely occurred during quenching and the black oxide film formed during tempering.

Of the 12 bolts that were tensile tested (described earlier), four were selected for SEM examination. The fracture surfaces showed a

mixed intergranular and dimple rupture topology and a fibrous texture similar to Zones 2 and 3 of the failed bolts. There was no evidence of a black oxide film.

Of the eight inventory and field bolts which were torqued to failure (described earlier), two were also examined in the SEM. Shear dimpling was prevalent, as expected in torque failures. The cup structure of shear dimpling was found throughout the fracture surfaces except for the 45° shear planes. All the bolts torque tested exceeded the minimum requirements for maximum torque-to-failure by 2 to 3 times.

Quench Cracks: Quench cracks in steel result from stresses produced during the austenite-to-martensite transformation, which is accompanied by an increase in volume. The observed cracks in the failed bolts meet the following characteristics of quench cracks:[*] the crack runs from the surface toward the center of mass, grows and exhibits a shear lip at the outer surface; the crack doesn't exhibit any decarburization in a microscopic examination; when tempering after quenching, the fracture surface is blackened by oxidation. Any condition that concentrates stresses that occur in quenching, promotes the formation of quench cracks. Distribution of mass and lack of uniform or concentric cooling of the part may promote cracking. In addition, selection of an unsuitable quenching medium may also be contributory. After quenching the part should be tempered as soon as possible to relieve the internal stresses formed in quenching (temper while the part is still warm, i.e., 150 to 200°F as withdrawn from the quenching medium).

Further support of a quench crack fracture mechanism may be found in examination of both light optical and electron microscopic fractographs of typical quench cracks in a 4340 steel.[+] These fractographs show the quench crack crescent where the crack is intergranular in nature. Comparable fractographs of the two failed bolts (Figures 4 and 5) show the same features; the quench crack crescent designated Zone 1, and the intergranular fracture mode in Zone 1. Fractographic examination of the bolt which failed magnetic particle inspection and tensile testing also support the contention that the cracks were pre-existing quench cracks and not due to the service environment.

Stress Corrosion Cracking, Hydrogen Embrittlement: High hardness steels can fracture under very low static stresses if they are embrittled by hydrogen absorption or exposed to an environment capable of causing stress corrosion cracking (SCC). Hydrogen embrittlement (HE) fractures frequently result from hydrogen permeation into a metal during electroplating and can be difficult to distinguish from SCC fractures, particularly when the environment is also a source of hydrogen. Both mechanisms usually result in faceted, intergranular fracture origins in low alloy steels.

Hydrogen produced during cadmium plating can lead to catastrophic failures of a stressed structural part. These plated parts must be baked to drive out hydrogen, lowering the internal concentration and thus reducing possibility of failure. Conventional bright cadmium deposited from cyanide baths is preferred due to its appearance and protective characteristics. However, this plating is a barrier to hydrogen diffusion,even prolonged baking may not drive off all of the hydrogen. The degree of embrittlement becomes more severe with increasing strength. For example, 4340 steel, 260 to 280 ksi UTS. with an acute notch (Kt =55.6) might be embrittled with less that 0.1 ppm mobile hydrogen. Therefore, low embrittlement baths are used for high strength steels. These produce duller and more porous plates which lose hydrogen more readily upon baking. However, the dull cadmium is not as protective as the bright.
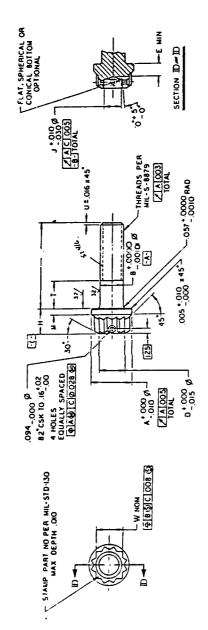
Specifications for baking cadmium-plated high strength steels to relieve hydrogen embrittlement tend to be vague, i.e., bake for 1 to 5 hours at 300 to 400°F. In aircraft applications, it is common practice to bake for 24 hours at 375°F for the highest strength steels, or use a sliding scale, depending on strength level. Although fractographic examination of the failed bolts showed intergranular fracture origins (Zone 1) which occur for HE and SCC, it is unlikely that the SCC or HE caused the failures. Both HE and SCC do not produce the black oxide observed. The only mechanism for this oxide is thermal growth during heat treatment. The stress durability test which was specifically designed to demonstrate effects of HE caused by electroplating or exposure to other environment containing a source of hydrogen showed no failures after stressing at 80 percent of the UTS for 200 to 400 hours. The preponderance of evidence attributes the failure of the two bolts during installation to the presence of pre-existing quench cracks.

Recommendations:

1. Insure 100 percent magnetic particle inspections of bolts after the tempering operation.

2. Specify dull cadmium plate to mitigate the potential for delayed failures due to HE or SCC.

3. Insure a 24-hour embrittlement relief baking at 375°F to remove and redistribute hydrogen within the bolt to prevent HE failures.

4. Alternatively, specify vacuum cadmium plate or ion-plated aluminum to eliminate the potential for hydrogen embrittlement.

5. Insure the radius at the shoulder/shank interface conforms to specification requirements.

6. Review the vendors fabrication operation on-site with technical experts (metallurgist, chemist) from the AMCCOM and ARL.

7. Replace all existing bolts with new or reinspected inventory bolts to mitigate the possibility of undetected small quench cracks growing under firing loads.

*Metals handbook, v.10, Failure Analysis and Prevention, 1975, p. 74.
+*Metals handbook*, v.9, Fractography and Atlas of Fractographs, 1974, p. 308.

Figure 1. Engineering Drawing of Bolt

PART NO	THREADS	A	B	D	E	H	J	L	M	T	U	W
9328588-1	500-20 UNJF-3A	.828	.4995	.562	.254	.504	.390	1.830	.123	.590	.047	.5625
9328588-2	500-20 UNJF-3A	.828	.4995	.562	.254	.504	.390	2.190	.123	.950	.047	.5625
9328588-3	5625-18 UNJF-3A	.938	.5615	.625	.287	.557	.440	1.830	.133	.590	.062	.6250
9328588-4	5625-18 UNJF-3A	.938	.5615	.625	.287	.557	.440	2.190	.133	.950	.062	.6250

ANSI Y14.5-1973 APPLIES.

MATERIAL:
STEEL, ALLOY 4140, SPEC MIL-S-5626 OPTIONAL
STEEL, ALLOY 4340, SPEC MIL-S-5000;
ALLOY 8740, SPEC MIL-S-6049;
ALLOY 6150, SPEC MIL-S-8503

THE BOLT SHALL MEET THE REQUIREMENTS OF SPEC MIL-B-8831 EXCEPT QUALIFICATION TESTING. QUALIFICATION TESTING IS THE RESPONSIBILITY OF THE QUALIFIED MANUFACTURER AND SHALL BE PERFORMED BY THE QUALIFIED MANUFACTURER THAT HAVE NOT DONE SO WITHIN THE LAST 3 YEARS.

STAMP PART NO PER MIL-STD-130 MAX DEPTH .010

SECTION ID—ID

FLAT, SPHERICAL OR CONICAL BOTTOM OPTIONAL

THREADS PER MIL-S-8879

Figure 3. Blacklight Photograph of Cracked Bolt MAG 1X.

ZONE 1
(BLACK OXIDE)

ZONE 2
(RIVER MARKS)

ZONE 3
(FIBROUS)

ZONE 4
(SHEAR LIP)

Figure 4. Bolt Head Showing Fracture Regions MAG. 2.5X.

Figure 2. Tempered Martensitic Microstructure MAG. 1KX.
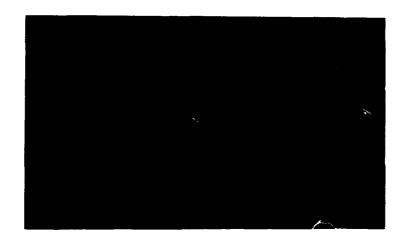
Figure 5. SEM Showing Intergranular Fracture MAG. 1KX.


```
Vert=    2073 counts   Disp= 1              Elapsed=    200 secs
                                                          UTW
                                            FE


  O



 FE  AL                              CR      FE
                                      MN
<- 0.320    Range=  10.230 keV                   10.230 ->
                                Integral 0  =    128550
```
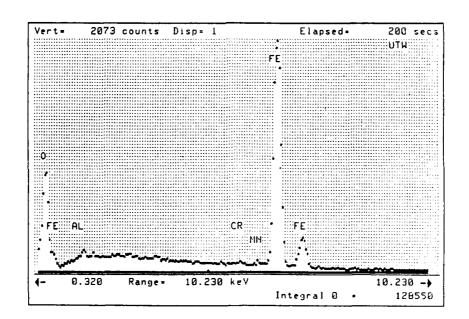
Figure 6. EDS Spectrum of Black Oxide

116

Figure 7. SEM of Mixed Mode of Fracture MAG. 1KX.
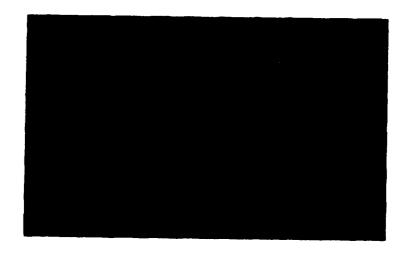
Figure 8. SEM of Shear Dimples, MAG 1.5KX.

117

PROACTIVE MAINTENANCE - THE NEW TECHNOLOGY FOR COST EFFICIENT CONTAMINATION CONTROL OF MECHANICAL MACHINERY

H.J. Borden
J.C. Fitch
J.W. Weckerly

Diagnetics, Inc.
5410 South 94th East Avenue South
Tulsa, Oklahoma 74145
(800) 788-9774

Abstract: It has been proven that almost all mechanical failures are caused by contamination; hard particle contamination to be specific. Once the root cause of machine failure has been defined, a program to correct these failures, extend machine life, and reduce maintenance costs must be developed. Such a program has been developed; it is called Proactive Maintenance.

Proactive maintenance is a three-step program that begins with the individual mechanical equipment and setting target cleanliness levels (benchmarks). The second phase deals with the system design, adequate filtration, and contamination exclusion techniques. The final step involves system monitoring. This process of continual monitoring is to ensure fluid and system cleanliness.

This paper is directed toward companies and manufacturers that have an interest in an efficient, cost effective maintenance program. To achieve total maintenance excellence, one must start at the beginning by taking an aggressive approach to maintenance technology.

Key Words: Abrasive wear; Contamination control; Contaminant monitoring; Fluid cleanliness; Machine life extension; Proactive maintenance; Root cause analysis.
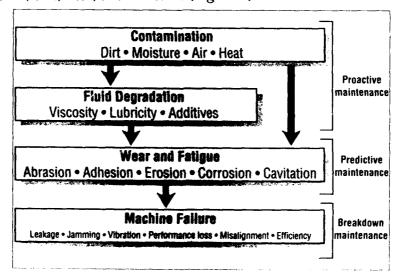
Introduction: Today, hydraulic and lubrication systems are being built more efficiently than before. Most hydraulic systems come equipped with filters as a standard and are not offered as an accessory. These better-made systems are by no means meant to last "forever." The theory of "buy it and leave it alone," tends to incur high contamination problems, downtime, and maintenance repair costs. Often times the blame for this short machine life is placed on faulty machine design, but the fault really lies with poor service and maintenance techniques.

119

These inadequate maintenance services fall into the category of *Breakdown maintenance*, which is essentially waiting for equipment to become inoperable before any maintenance is observed. Another form of ineffective maintenance is *Preventive maintenance*. This maintenance philosophy is dependent upon a specific date or number of cycles and the availability of money and/or maintenance personnel. *Predictive maintenance* is a more current form of maintenance that uses non-destructive instruments to help predict a failure that is already in progress. This maintenance is less than optimal, because a failure has already begun. Using predictive maintenance, this failure will not lead into a catastrophic breakdown, but there will be maintenance costs, downtime, and production loss.

A new age of maintenance philosophy has come about in the 90's. The philosophy of maintaining higher fluid cleanliness levels, extending machine life, and defining the root causes of failure. This *Proactive maintenance* philosophy needs to be adopted for companies and manufacturing firms to achieve total quality and cost effective maintenance. Proactive maintenance is aimed at identifying and correcting failure root causes, extending individual machine life, and reducing maintenance costs. This can be achieved through a simple three-phase strategy listed below:
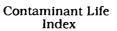
Phase One: The first phase begins with the training and understanding of proactive maintenance and its goals. Proactive maintenance is a condition-based maintenance strategy, as such maintenance is dependent upon the real-time needs of the machine. Maintenance is prescribed when changes occur to specific operating conditions (failure root causes), and these changes present a risk to a machine's health.
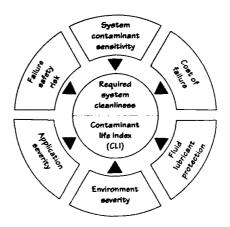
One of the main conditions that present a great risk to a machine's operating health is excessive contamination. There are four types of contamination that are dangerous to any machine's operational life, and they are air, dirt, heat, and moisture (Figure 1).

Contamination
Dirt • Moisture • Air • Heat

Fluid Degradation
Viscosity • Lubricity • Additives

Proactive maintenance

Wear and Fatigue
Abrasion • Adhesion • Erosion • Corrosion • Cavitation

Predictive maintenance

Machine Failure
Leakage • Jamming • Vibration • Performance loss • Misalignment • Efficiency

Breakdown maintenance

These contaminants are easily classified as root causes of failure. The first phase of proactive maintenance is to identify and correct the main failure root causes in a machine. Dr. Leonard Bench of Pall Corporation, states that 70% - 85% of all mechanical failures are caused by hard particle contamination and 90% of these failures are caused by abrasive wear. A recent report published by Lubricant Engineering magazine leads to the conclusion that more than 82% of wear related losses are contaminant induced. Notice that the largest portion of this is abrasive wear. From these findings, it would be advantageous to concentrate maintenance activities on correcting hard particle contamination, which causes 82% or more of the mechanical failures, than spreading the maintenance time out between three other root causes which would only eliminate 18% or less of the breakdowns.

Now that hard particle contamination has been defined as the root cause of failure, something must be done to correct it. Phase one consists of the setting up benchmarks for each individual machine. These benchmarks are actually goals; fluid cleanliness level goals for individual pieces of equipment. To have a condition-based maintenance program, one must know the current condition of the machines and also have a known benchmark that is to be achieved. This known target cleanliness level is extremely important. A fluid cleanliness benchmark must be set according to each individual machine. The Contaminant Life Index (CLI) is a simple method to achieve this benchmark. The CLI is a set of ten questions based on the factors that can influence a machine's cleanliness level needs (Figure 2).

Contaminant Life
Index

Another method to identify a cleanliness benchmark is to use the Life Extension Method (LEM). This method uses the aid of three different tables. The appropriate table is selected to match the machine type. The benchmark is represented in International Standard Organization - ISO Code.

Phase Two: Once a cleanliness benchmark has been obtained, the next phase is to achieve and maintain that goal. Phase two is mostly dependent upon proper filtration and contaminant exclusion techniques. Before filtration needs are specified, exclusion techniques must be discussed. Initially, it is less costly to keep the contaminants out of the fluid altogether, than to remove them once they are in the fluid. The first step to contaminant exclusion is to identify the sources of contaminant ingression and then correct them. For hydraulic equipment, cylinder wiper seals is the most common entry point. The best way to combat this ingression is to use boot seals and good wiper seals. Unnecessary component repair and replacement is another source of contaminant ingression. Try not to open up any sealed components if possible and when repair is necessary, flush and clean the components at low levels before putting them back into service. This flushing technique is also good for getting rid of built-in contaminants in new equipment. New oil is a large source of contaminant ingression. Keep the fluid suppliers honest by checking their new oil cleanliness.

Proper filtration means the accurate selection, location, and installation, or upgrading of current filtration, to achieve the aforementioned cleanliness benchmark. The filter selection must be application, environment, and machine specific. This can be accomplished through the use of a Filter Selection Chart (FSC). This methodical means of filter selection consists of questions somewhat like the CLI. After answering the questions and doing a little math, the FSC will identify the proper filter to be installed. Some companies depend upon a filter salesperson to supply this information, but more often than not, the salesperson is not equipped to select filters objectively. One other fact that is often overlooked when dealing with filters is tank breathers. High efficiency breather filters should be used on tanks and reservoirs.

Phase Three: The final and maybe the most important phase in implementing proactive maintenance is to set a rigorous contaminant monitoring schedule. This contaminant monitoring technique is critical to effective contamination control. The control is achieved by monitoring the individual machines providing regular feedback on contaminant levels.. Maintenance personnel are able to check and insure that the cleanliness benchmarks are being maintained and that the filters are operating properly. Continual monitoring allows for the condition and health of any machine to be known, present or past. Continual contaminant monitoring has proven to be cost efficient, because the operating life of a machine is actually extended since it is not allowed to progress towards failure.

Proactive Maintenance: The condition-based philosophy of proactive maintenance meets the objectives of identifying and correcting failure root causes, extending machine operation life, and reducing maintenance repair costs. The three phases: 1) Setting benchmark cleanliness levels, 2) Selecting and installing proper filtration, and 3) Monitoring fluid contaminant levels; are cyclical and must all be implemented at the same

time. These three phases prove to be very dependent upon each other if a truly cost efficient contamination control program is to succeed.

When a total quality and cost effective maintenance program is being considered, such as proactive maintenance, a total turn-key installation should be considered. This proactive philosophy must be used at all times during the training, installation, and assignment of field personnel to the job of maintaining hydraulic and lubrication machines. An Installed Proactive Maintenance Program (IPMP) is the most timely and cost efficient strategy to use in contamination control.

Bibliography:

Bensch, Dr. Leonard E., "A Modern Review of Field Contamination Levels Based on Analysis of 25,000 Samples," International Fluid Power Applications Conference, March 1992, Paper No. I92-15.6.

Borden, Holly and James C. Fitch, "Strategic Implementation and Cost/Benefit of Contamination Control," International Fluid Power Applications Conference, March 1992, Paper No. I92-15.7.

Fitch, E.C., Fluid Contamination Control, Stillwater: FES, Inc., 1988.

Fitch, James C. and Holly J. Borden, "Interpreting Contaminant Analysis Trends into a Proactive and Predictive Maintenance Strategy," British Hydromechanics Research Association, December 1992.

Fitch, James C., "Proactive Maintenance: The Buzz Word for the 90s," Utility Construction & Maintenance, September/October 1991, pp. 14-17.

Fitch, J.C., "Quantifying the Contaminant Tolerance of Hydraulic Systems Using the Contaminant Life Index," National Conference on Fluid Power, 1986.

Mathew, J., and J.S. Stecki, "Comparison of Vibration and Direct Reading Ferrographic Techniques in Application to High-Speed Gears Operating Under Steady and Varying Load Conditions," Lubrication Engineering, August 1987.

Pilot Survival Rubber Dinghy (Liferaft) - Leaking of CO_2-High Pressure Gas Cylinders

K. Wolf and W. Schoenherr
Wehrwissenschaftliches Institut fuer Materialuntersuchungen (WIM)
Landshuter Str. 70,
8058 Erding, Germany

ABSTRACT

CO2-high pressure gas cylinders were checked for leakages after rejection. Two of them were empty. One empty gas cylinder was investigated at WIM. Leakages were proved to exist around the cylinder heads after pressure tests in water bath using varying tightening torque values of the valve bodies and tappet washer insets. After exposure of the cylinder heads it was evident that the sealing washers showed slight inequalities caused during the production process, the contact surfaces showed deep tool marks and were too small. The valve body material was not anodized on the contact surfaces and, accordingly, was corroded (crevice corrosion), as a result of material consolidation owing to the tightening torque applied the sealing effect of the sealing elements was destroyed after some time.
In order to avoid such defects, which may take fatal effect on the aircrew, the urgent advice was given to modify the construction of the cylinder heads.

Key: Cylinder, sealing elements, crevice corrosion, Al-alloy

INTRODUCTION

After a pilot's bailout above sea, the pilot survival rubber dinghy (liferaft) is filled automatically trough a CO2 cylinder. In our case CO2 cylinders were checked and two were found to be empty. In an emergency case this would have meant death to the crew. Now the question had to be answered why the cylinders had been empty:"Had the CO2 cylinders not been filled or were they leaking?" In order to prove that the cylinders had been properly filled and determine the location of any leakage, the cylinders were subjected to a pressure test in a water bath. By applying various pressure methods and various tightening torque values, leakages were proved to exist near the cylinder connections. After this had been found out, the cylinder heads were disassembled for the purpose of understanding the set-up and the functioning of these heads and of determining the component which caused the leakage. Near the sealing elements

Dr.W.Schoenherr, Department Head, WIM
K.Wolf, mechanical engineer, WIM

a number of design and fabrication characteristics were found to promote damage. A design modification of the cylinder heads is absolutely necessary according to these findings.

RESULTS

Material:

- A completely assembled CO_2 cylinder with valve body and tappet washer inset (damaged part - sample 1)
- A cylinder with loosened valve body and new tapped washer inset not yet subjected to pressure (damage part - sample 2)
- New tappet washer inset that has not been subjected to tightening torque or pressure (reference part - sample 3)
- Part of a tappet washer inset, designated "E" that has been subjected to tightening torque and pressure (reference part - sample 4)

Pressure Test:

The bottom part of the cylinder No.1 was provided with a nipple and filled with air at an overpressure of 10 bars. Air escaped at the lateral borehole of the cap thread (Figure 1 and 2). A set-up of the cylinder head with valve body and tapped washer inset is outlined in Figure 3. The head of cylinder 2 was screwed off and connected to a pressurized-air pipe through a fitting part <1, 2>. The tapped washer inset was tightened in the valve body with a torque of 13.5Nm in the first test and 15.7Nm in the second test. The manufacturer prescribed a torque value of 13.5 to 15.7Nm. For both tests, an overpressure of 80 bars was set. The operational pressure was stated to be 56 to 60 bars. Under these conditions the system showed no leakages for a period of 20 min.
A test series with varying tightening torques, which were applied at different internal pressures (test run A with 80 bars internal pressure and test run B with no set internal pressure), was carried out to show when the system shows leakages <3>. The test conditions were: test pressure 80 bars, test medium nitrogen, tested object under water, tightening torques increasing from 7.8 Nm to 15.7 Nm in 2-Nm steps and afterwards falling equally. The results are shown in the following table:

Test run A		Test run B	
Tightening torque	Sealing effect	Tightening torque	Sealing effect
9.8 Nm	leaking	7.8 Nm	leaking
11.8 Nm	leaking	9.8 Nm	leaking
13.7 Nm	leaking	11.8 Nm	not leaking
15.7 Nm	not leaking	9.8 Nm	not leaking
13.7 Nm	not leaking	7.8 Nm	not leaking
11.8 Nm	leaking	5.9 Nm	leaking

Test run A and test run B differ as to the tightening torque that provided a tight seal. For each test object, test run A was carried out to be followed by test run B; this leads to the following considerations in terms of tightening torque deviations:

- Tightening under test pressure requires higher tightening torque values.
- After test run A, the sealings may have adjusted to at least one sealing surface, thus requiring a lower tightening torque.
- The coefficient of friction was probably reduced after test run A; this also reduced the required tightening torque.

The test runs have shown that the increasing tightening torques of 15.7 Nm are necessary because the original specified tightening torques were not enough to ensure a tight seal with a safety margin.

Fractographic and Surface Examination

The whole valve body 1 but not the contact surface was given an anodic coating of approx. 2 μm. When installed, the valve body is connected with the surrounding atmosphere through a relief well. The entire circular contact surface of the valve body was locally corroded (Figure 4). The Figures 5, 6, 7 and 8 show clearly that it is crevice corrosion. The corrosion started from the outside of the valve body. On the contact surface and in the cavity next to the contact surface glycerin was found; glycerin is hygroscopic. Saliferous condensation introduced with humid saliferous air via the relief well accelerated the corrosion process. Due to the interconnected and deep corrosion, it was not possible to seal the spot by means of a sealing washer. The tappet washer C, which had been subjected to torque and pressure, was examined with a scanning electron microscope: there were no indications of corrosion or crack initiation. The paint coat of the tappet washer was engraved in the surface by the tightening torque or pressure impact.

Metallography

The tappet washer inset that had not been subjected to torque or pressure (sample 3) was opened. The tappet washer was embedded for grinding. It was even and showed no plastic deformation caused by the identification signs. The thickness of the painted signs was 11 to 13 μm (Figures 9 and 10). The tappet washer inset of sample 2 which had been subjected to pressure due to the above tests was also exposed in order to determine the condition of the contact surface and the tappet washer. Figure 11 shows the tappet washer surface designated "A". The interior side of the washer showed the sign "A" convexly (Figure 12). The tappet washer was deformed, leakage was caused (Figures 9, 10, 11 and 12). Figures 11 and 12 show clearly that a single identification mark may be as large as the entire surface width, so that leakages may be expected even without any corrosion. The contact surface of the sealing washer showed no engravings because the pressure was exerted only for a short time. This led to the conclusion that the deformation was

127

merely elastic. The sign "E" that was engraved in the tappet washer of sample 1 and sample 4 showed a similar appearance in both samples. The sealing washer in sample 1, which had been installed for three years, was taken out and showed only slight engravings. This indicates a reduction of the sealing material's elasticity. By way of a cross section it was shown that the valve body is a forged piece which has been machined. In the area next to the contact surface, tuberculation was visible with a depth of up to 60 μm (Figures 13 and 14). The contact surface was corroded as deep as 50 percent of the wall thickness (Figures 15 and 16).

Chemical Investigation

According to the results found by chemical investigation, the valve body material is AlMgSi 1 (Table 1). The sealing washer is made by fiber-reinforced phenolic resin.

Table 1: Chemical composition

Samples or reference material	Percentage of elements by mass							
	Si	Fe	Cu	Mn	Mg	Cr	Zn	Ti
Valve body of sample 2	1.16	0.21	0.02	0.50	0.70	0.01	0.03	0.01
AlMgSi 1 DIN 1725	0.70 to 1.30	$\leq$ 0.50	$\leq$ 0.10	0.40 to 1.30	0.60 to 1.20	$\leq$ 0.25	$\leq$ 0.20	$\leq$ 0.10

CONCLUSION

The examined CO_2 cylinders for the liferaft of the MRCA weapon system were leaking at the valve bodies and the contact surfaces of the tappet washer inset in the cylinder head. Besides that they showed weak points which can lead to leakage at other places. The reasons for these deficiencies were found out and have been summarized as follows:

- Contact surface 1 was uneven due to the engraved paint coat.

- Leakage of contact surface 2 is possible because after longer-lasting pressure on the system, the sealing material showed a reduction of elasticity;

- Sealing surface 5 had not been anodized; therefore it corroded, which again caused leakages. The corrosion was the crevice corrosion type;

- Another deficiency was the small contact surface (5) of the valve body, which showed deep tool marks and was considered a design deficiency;

- Under the above mentioned conditions the tightening torque values prescribed had been chosen too low.

To make sure that such life-threatening design deficiencies will definitely never occur again, the urgent advice was given to modify the design of the cylinder head with valve body and tappet washer inset.

SUMMARY

The examination led to the result that several inappropriate constructions and deficiencies caused by inadequate production methods in the sealing system of a CO2 cylinder may cause life-threatening situations.

ACKNOWLEDGEMENTS

REFERENCES

1. K.Nagel, K.O.Cavalar: Pruefung von Stahlflaschen auf Herstellungsfehler und Betriebsschaedigungen, DGZfP-Jahrestagung 1967

2. W.Woerlen: Neue EWG-Einzelrichtlinien fuer Druckgasflaschen. TUEV, Bd. 27 (1986), Nr.4

3. M.Baumgaertner, H.Kaesche: The nature of crevice corrosion of aluminum in chloride solutions, Werkstoffe u. Korrosion 39, 129-135 (1988)
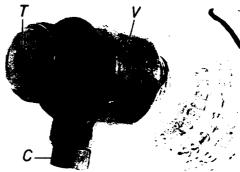
Figure 1   1:3
CO2 Cylinder with valve body (V)
and tappet washer inset (T)

Figure 2   1.3:1
Valve body (V) with adjusted tap
pet washer inset (T), thread cap
(C) and drilled hole
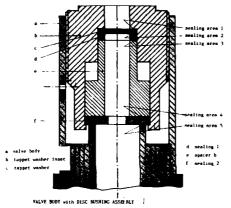
Figure 3
Schematic representation of the
valve body with tappet washer
inset
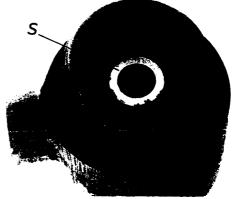
Figure 4   2.1:1
Valve body without tappet washer
inset; corrosive attacked con-
tact surface (S)

Figure 5
The SEM-photograph shows the cor-
rosive attacked and poorly manu-
factured contact surface 5 (S)

Figure 6
The contact surface 5 (S) and
the adjacent surface (AS) were
attacked by crevice corrosion

130
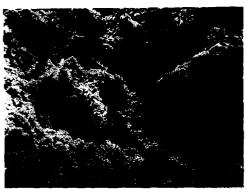
Figure 7
Deep cracks and corrosively
attacked grains characterized
the contact surface

Figure 8
The cylindric shaped surface
adjacent to areas of the con-
tact surface 5

Figure 9 126:1
Cross section:the tappet washer
of specimen 3 without deforma-
tion around the designation "A"

Figure 10 1 000:1
The thickness of designation "A"
was 12 to 13 μm

Figure 11 7:1
The designation of the extended
to the entire width of the con-
tact surface (specimen 2)

Figure 12 7:1
Interior side of the tappet was-
her of specimen 2 with the push-
ed paint designation
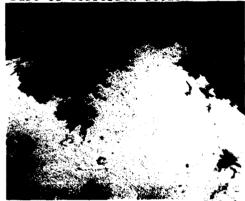
Figure 13 3.5:1
Cross section of the valve body;
the forged part was heat treated
and than manufactured with tools

Figure 14 63:1
Section of Figure 13; the wall
(W) and the transition to the
contact surface (S) are scare-
marked with 60 μm holes as a re-
sult of corrosion attack

Figure 15 20:1
Cross section parallel to the
contact surface; the corrosion
attack extended nearly to the
entire thickness of the wall

Figure 16 500:1
Section of Figure 15; corrosion
attack damaged widely the con-
tact surface

132

A DEDICATED COMPRESSOR MONITORING SYSTEM EMPLOYING CURRENT SIGNATURE ANALYSIS

K. N. Castleberry and S. F. Smith
Oak Ridge National Laboratory*
Oak Ridge, Tennessee 37831-6006

Abstract: The use of motor current signature analysis (CSA) has been established as a useful method for periodic monitoring of electrically driven equipment. CSA is, moreover, especially well suited as the basis for a dedicated continuous monitoring system in an industrial setting. This paper presents just such an application that has been developed and installed in the U.S. government uranium enrichment plant at Portsmouth, Ohio. The system, which is designed to detect specific axial-flow compressor problems in 1700-hp gaseous diffusion compressors, is described in detail along with an explanation of detected fault conditions and the required signal manipulations. Amplitude demodulation and subsequent digital processing of motor signals sensed from area control room ammeter loops are used to accomplish the desired monitoring task. Using modified off-the-shelf multiplexing equipment, a 386-type personal computer, and special digital signal processing hardware, the system is presently configured to monitor ten compressors but is expandable to monitor more than 100. Within its first few days of operation in September 1992, the system detected a compressor problem that, when corrected, resulted in a cost avoidance of about $150,000, which more than paid for the hardware and software development costs. Finally, plans to expand system coverage in the coming year are also discussed.

Key Words: Demodulation; amplitude demodulation; remote sensing; compressor monitoring; rotating stall, current signature analysis.

Introduction: Since 1987, personnel from the Instrumentation and Controls Division of the Oak Ridge National Laboratory (ORNL) have been investigating and developing motor current signature analysis techniques for identifying problems and abnormal operating conditions in electric-motor-driven equipment.[1] CSA has been shown to be useful in the diagnosis of conditions such as rotor imbalance, coupling misalignment, compressor cavitation, surging or rotating stall, fan and pump drive-belt damage, and other problems not normally thought to be observable by examination of the motor current.[2] Moreover, many load-related problems have been found to be more easily detected with CSA than with any other single sensor means. Part of the ORNL work has involved studies of the motors and compressors used in the U.S. Department of Energy's uranium enrichment facilities at Portsmouth, Ohio, and Paducah, Kentucky. These plants use a process known as gaseous diffusion to enrich uranium in the form of uranium hexafluoride and employ a cascade of many hundreds of compressor stages driven by motors ranging in size from 100 to 3300 hp. The second largest size of compressors used are driven by 1750-hp motors and are referred to as 00-size (or just 00) compressors. Each of these axial-flow compressors contains over 1000 blades, which range in length from about 3 to 8 inches. Five hundred 00 compressors arranged in cells (groups of ten) are contained in about half of the X-330 building at the Portsmouth Gaseous Diffusion Plant (GDP).

From 1974 though 1991, fifty eight OO compressors failed in the X-330 building because of what are listed in the plant compressor failure data base as unknown causes. From FY 1989 through FY 1991, 16 such failures were recorded, with seven of these located in the stage-one positions of cells. Rotating stall is suspected to be the primary cause of most OO stage-one failures because it can go undetected for days at a time. The length of time that a compressor can operate in rotating stall varies with operating power, but usually within a few days the cumulative stress will deblade the compressor. When a compressor does fail, the entire cell of ten compressors must be bypassed, taken off-stream, and shut down, sometimes for several weeks, to allow damaged components to be replaced. Besides the incurred maintenance, costs there are losses in both cascade efficiency and separative work capacity.

Surge and Rotating Stall: Figure one shows a typical mechanical configuration and operating characteristic for a GDP compressor. A normal operating point would fall somewhere on the characteristic curve and would be determined by the system in which the compressor is
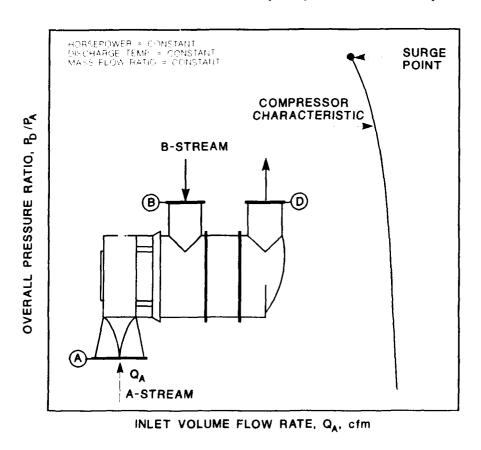
Fig. 1. Typical GDP compressor characteristic

installed. As system restrictions increase, the operating point moves up the curve to a higher compression ratio and a slightly lower volume flow. Further increasing restrictions drives the compressor to the surge point, where the operation of the compressor becomes unstable. This instability takes one of two forms, surge or rotating stall, and both conditions

134

involve the existence of what is called the secondary operating characteristic of the compressor (Fig. 2).
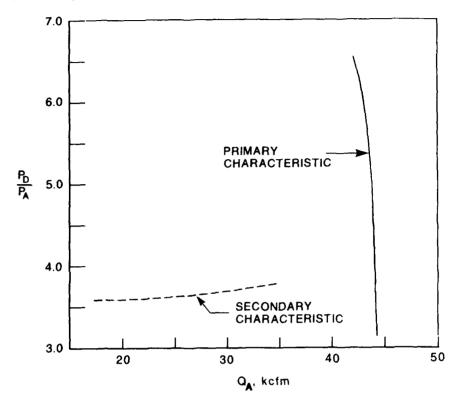
Fig. 2. Primary and secondary operating characteristics.

The difference between surge and rotating stall is illustrated in Fig. 3. Surge is a large-amplitude oscillation of the flow though the compressor which involves repeatedly moving the instantaneous operating point from the primary to the secondary characteristic and back.[3] It is usually easy to detect because of the resulting motor ammeter fluctuations or the distinctive sound made by the compressor. Rotating stall, on the other hand, is much more subtle and is characterized by the formation of a bubble-shaped region of recirculating gas that rotates within the compressor. This region, which is usually called a stall cell, effectively blocks a portion of the cross-sectional area of the compressor and results in a decrease in both efficiency and compression ratio. When forced into this abnormal mode of operation, the compressor operating point moves from its primary characteristic to a point on its secondary characteristic. In many industrial systems, when a compressor moves into rotating stall, it will remain there until the compressor fails or until operator intervention restores normal system flow. Since rotating stall results from operation of the compressor on the secondary characteristic, it is sometimes referred to as secondary stable operation or just secondary.

Secondary operation increases the risk of compressor failure by increasing vibration levels, which increase blade temperatures and other internal mechanical stresses. The amplitude of vibratory stress in the blades during secondary can be five times the level that occurs

135

PRESSURE RATIO

INITIAL TRANSIENT MOTION OF OPERATING POINT

S

INLET VOLUME FLOW RATE

ROTATING STALL

REGION OF NORMAL FLOW

ROTATING REGION OF LOW FLOW

CIRCUMFERENTIALLY NONUNIFORM FLOW

AXIS OF ROTATION

COMPRESSOR ANNULUS

LARGE AMPLITUDE AXIALLY OSCILLATING FLOW

PRIMARY CHARACTERISTIC

OSCILLATORY MOTION OF OPERATING POINT DURING SURGE

SECONDARY CHARACTERISTIC

PRESSURE RATIO

Q_{min}

INLET VOLUME FLOW RATE, Q

SURGE

Fig. 3. Differences between surge and rotating stall.

during normal operation. Over time, the resulting mechanical stresses can fatigue internal parts, especially blades, and result in compressor failure. Secondary is known to occur in many types of axial flow compressors besides those used in the GDPs. In other types of industrial systems where compressors operate alone or in small groups, secondary may not be difficult to detect, but resulting flow upsets may seriously disrupt normal system operation. Some jet aircraft engines, for instance, are known to occasionally experience secondary operation. When this occurs, it is immediately evident, but the associated drop in power output can have devastating consequences, especially when it occurs during a period of high power demand such as takeoff.

Although secondary operation can, over a period of time, deblade a compressor, it can be easily missed or mistaken as normal especially in the GDPs where there are hundreds of compressors to consider. Any unusual sounds that a compressor might make as a result of a rotating stall cell are often masked by the sounds from surrounding equipment. A low compression ratio, which is characteristic of secondary, is not a foolproof indicator because it can also result from operation on a low compression ratio part of the normal operating curve. Several years ago it was determined that the presence of the stall-cell passing frequency in the vibration spectrum of the compressor was perhaps the most conclusive evidence of the presence of rotating stall. Tests in many types of axial-flow compressors have shown that a stall cell typically rotates at slightly less than half of running speed or about 13 Hz in the case of a GDP 00 compressor running at 1800 rpm. This vibration can often be seen by an accelerometer mounted on the motor-end bearing housing of the compressor (Fig. 4). A 13-Hz shaft displacement can also sometimes be seen by a
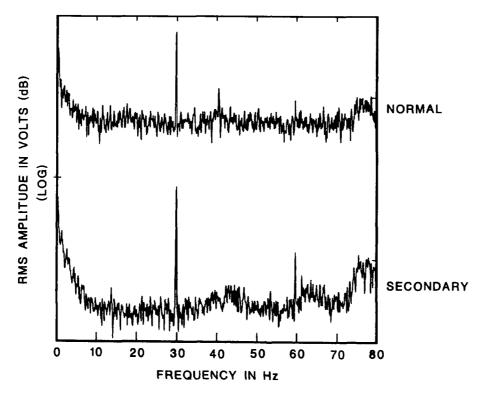
Fig. 4. Spectra of load-bearing, vertical-axis accelerometer signals

137

displacement probe on the motor coupling (Fig. 5). The same stall-cell frequency is found to be much more apparent in a plot of the amplitude-demodulated motor current spectrum (Fig. 6). The detection of this vibration frequency with the accelerometers installed several years ago in the GDPs as part of the cascade automatic data processing systems was never viewed as very practical because of several concerns, which included accelerometer reliability problems, insufficient computing capability in the system, and marginal sensitivity.

Monitoring Considerations: It is a common industrial practice to feed process monitoring instrumentation signals to control rooms where areas of a given plant are monitored and controlled. The GDPs are no exception to this, employing area control rooms (ACRs) that oversee about 200 to 300 stages each. Since part of the standard stage monitoring instrumentation located in the ACR is a motor ammeter for each stage, these signals may be accessed and analyzed without installing special sensors on cascade equipment and running long signal lines to them. Thus, motor and compressor monitoring can be implemented for many stages distributed over a large area for a relatively low cost.

Monitoring System Overview: The prototype secondary monitoring system was designed to detect secondary operation in any of ten compressors by sequentially monitoring the motor current signal from each compressor stage. As each raw current signal is selected through a signal multiplexer, it is demodulated and sampled for about five seconds and then processed digitally to calculate the Fourier transform of the demodulated signal. The transform is then examined for evidence of a significant component between 12.5 and 13.25 Hz, and if it is found, the stage is resampled and checked again. The 13-Hz component must be found in two successive sample windows before the system will alarm and indicate secondary for the stage.

During scanning and sampling, the system graphically displays a screen of information in one of several formats. The display selection is operator controlled and can show status information for all ten monitored stages or the Fourier transform or time-data plot from the previously sampled stage. The system software will occasionally adjust the gain of a particular channel to optimize the signal level into the analog-to-digital (A/D) converter. These gain changes will be reflected, as needed, in the display scale.

Hardware: The X-330 Secondary Monitor is built around a typical 20-MHz, 386-type personal computer (PC) with enhanced color graphics capabilities (Fig. 7). The PC is installed in ACR-2 in the X-330 building at Portsmouth where it acts as both the controller and the operator interface for the system. The PC is supported by an external signal multiplexer (MUX) and A/D conversion hardware, which is located in the basement under ACR-2. The MUX is a Keithley WorkHorse system that contains a type AIN-16, 16-channel analog-input card, although it can accommodate up to seven input or output boards. In the prototype system, only ten of the available 16 input channels on the AIN-16 card are used. A parallel communications link between the PC and the MUX rack allows control signals and data to be transferred back and forth at a rate of up to 500 kilobytes per second. This link uses a Keithley WH-CIB-PAR board in the MUX rack and a WH-PCDB-PAR board in the PC.

The motor current signals are sensed directly from the control room ammeter loops using clamp-on current transformers (CTs) like the one pictured in the lower left corner of Fig. 7. These CTs are Fermitech Model 4LN2-5-333, which produce an output voltage of 333 mV rms for a full-scale current of five amps in the meter loop. The signal from each CT is run via a shielded, twisted-pair cable (Belden type 8762) to the MUX in the basement.
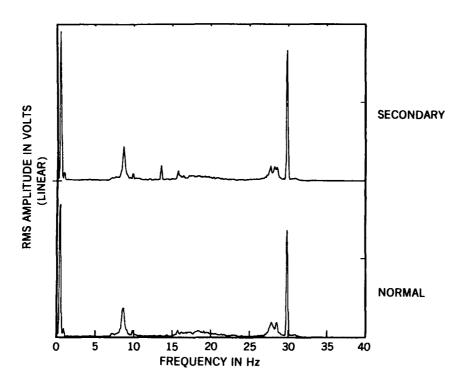
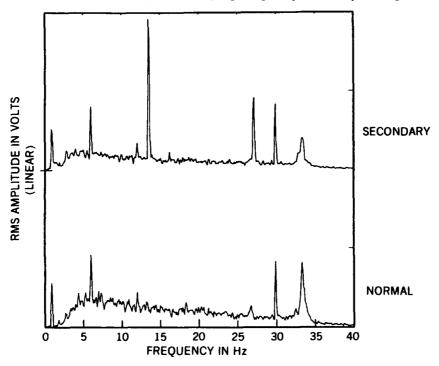Fig. 5. Spectra of compressor coupling flange displacement probe signals.

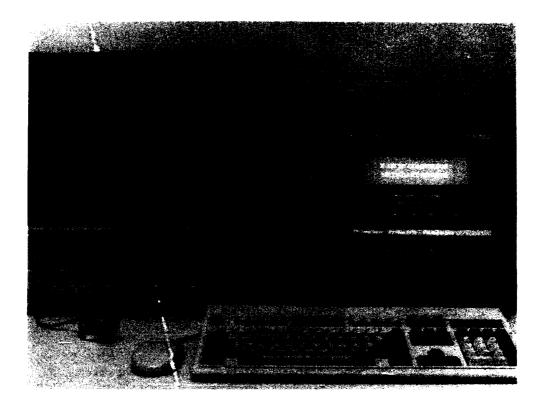Fig. 6. AM demodulated motor-current spectra.

139

Fig. 7. Secondary monitor hardware.

At the MUX the signals are sequentially selected, demodulated, and amplified. A special in-line AM demodulator was designed and added to the circuitry of the AIN-16 card between the input signal multiplexing and the on-board differential amplifier. The signal processing sequence implemented on the AM demodulator board is shown in Fig. 8. In this application, phase modulation rejection was not a problem so a precision full-wave rectifier provided adequate results as the demodulator stage. An integrated-circuit switched-capacitor filter, a MAX291, configured as an 8-pole, 30-Hz low pass, was initially placed immediately following the demodulator stage. This arrangement resulted in the presence of clock-intermodulation components as large as the 13-Hz component of interest in the output
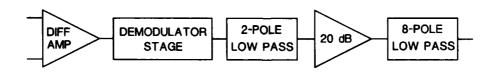
Fig. 8. AM demodulator implementation.

spectrum. To overcome this problem, an analog, two-pole, 30-Hz low-pass filter was placed

140

after the demodulator to attenuate the dominant 120-Hz signal and permit a 20-dB gain stage to be placed before the MAX291. This pushed the spurious signals from the MAX291 down into the noise floor and provided a satisfactory output signal. After the 8-pole filter, the signal is returned to the AIN-16 board, where it is sampled via the on-board 12-bit A/D converter. In just over five seconds 512 samples are taken at a rate of 100 samples per second under the control of the PC. Timing of the sample interval is provided by a timer board (Model DCC5 from Industrial Computer Source) in the PC, which interrupts the PC via the number-two interrupt request line 100 times a second. The interrupt service routine in the PC sends the sample command to the MUX.

When demodulating a carrier of frequency fc, the bandwidth of the extracted modulation can cover a maximum frequency range of $fc/2$ without experiencing potentially serious frequency aliasing. Therefore, for a 60-Hz carrier, only the band from 0 to 30-Hz is normally examined after demodulation. A 60-Hz sample rate would, according to Nyquist theory, provide the necessary sample timing to recover up to 30-Hz, but in practice for windowed sampling the sample rate should be 2.5 to 3 times the desired maximum frequency. The 100-Hz rate was chosen as a convenient figure that satisfies this criterion. Calculation of a floating-point Fourier transform from the sampled data is a fairly math intensive operation and could require several seconds if done by the PC alone. The time required for this calculation is reduced to milliseconds by passing the sample data to a digital signal processing (DSP) board (a TMS320C30 board made by Sonitech Inc.) and allowing it to calculate the transform. The transform yields a 256-point magnitude array covering a 50-Hz band, but only the lower 30 Hz of the transform data is actually used.

It should be noted that since the sampling operation uses most of the approximately five seconds for processing each signal, parallel sampling (sampling of two or more channels simultaneously) could provide a significant increase in system speed. This is certainly within the capabilities of the described computing and communication hardware since during the sampling intervals the PC central processor unit is essentially idle. Parallel sampling was not used in the prototype system because it requires the use of multiple A/D cards, but it will be employed in a planned expansion of the system.

Application Experience and Plans: In September 1992 the prototype system, designed to initially monitor ten 00 stage-one compressors, was installed in the X-330 building at Portsmouth. Immediately after installation and power-up of the system, it began to indicate that the compressor in stage 4.1.1 was operating in light secondary. The stage was being run with the recycle valve partially open, and all temperature and pressure readings for the stage appeared to be in the normal range. A partially open recycle valve, in this case, increases the compressor inlet flow and tends to suppress secondary so it was probably a precautionary measure since this stage had apparently been a problem in the past. According to the compressor failure data base, four compressors had failed in this location since November 27, 1989, with secondary as a suspected cause in each case. After noting the monitoring system alarm indication, building operations personnel opened the recycle valve further. This action removed any indication of a secondary stall cell from the motor current signal and probably prevented yet another failure.

An FY 1993 project is under way which will expand the secondary detection system to a capacity of 50 channels so that all of the stage-one 00 compressors in X-330 can be monitored. If the secondary monitoring system is successful in providing reliable compressor status information to X-330 personnel, it could virtually eliminate this failure mode in monitored compressors. With an average stage-one failure rate of 2.6 compressors per

141

year over the past three years and estimated compressor rebuild and change-out costs of around $150K per event, a system monitoring all 50 stage-one compressors in X-330 could provide a cost avoidance of about $390K per year. For the monitoring system, this would represent a total investment payback period of less than one year. Additional benefits in separative work would also be realized because of fewer off-line cells. Since secondary has also been known to occur in stage locations other than stage-ones, consideration is also being given to eventually expanding the system to monitor all 500 00-size compressors in the X-330 building. Installation of a similar system at the Paducah GDP is also a possibility in the near future.

References

1. S. F. Smith, K. N. Castleberry, and C. H. Nowlin, "Machine Monitoring Via Motor-Current Demodulation Techniques", *Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group*, Virginia Beach, VA., April 3-5, 1990, pp. 87 96.

2. S. F. Smith and K. N. Castleberry, "Advanced Techniques in Current Signature Analysis", *Proceedings of the 46th Meeting of the Mechanical Failures Prevention Group*, Virginia Beach, VA., April 7-9, 1992, pp. 263 275.

3. E. M. Greitzer, "The Stability of Pumping Systems - the 1980 Freeman Scholar Lecture", *Transactions of the ASME The Journal of Fluids Engineering* Vol. 103, pp. 193 242 (June 1981).

NONDESTRUCTIVE EVALUATION
AND
INFORMATION PROCESSING I

Cochairmen: Carol Lebowitz
              Naval Surface Warfare Center

              Richard Hayden
              Technology Integration, Inc.

A NEURAL NETWORK APPROACH TO WELD SYSTEM DIAGNOSTICS: AN OPTIMIZATION STUDY

Joel M. Milano   Lance A. Flitter   Richard Morris
Carderock Division, Naval Surface Warfare Center
Bethesda, MD 20084-5000

Abstract: The focus of this ongoing study is the development of a generic methodology, utilizing non-parametric and parametric statistical techniques, capable of extracting learned rules, feature maps, and other information from trained neural networks. This paper concentrates on the non-parametric analysis results for a neural network trained for weld acoustic monitoring. Information from both the weld acoustic data, as well as the neural network itself, was used in the optimization of the weld acoustic model, the learning model, and the neural network's physical design. Non-parametric analyses resulted in: (1) a simplification of the neural network from three hidden layers to one, with an associated reduction in processing time; (2) an increase in overall accuracy; (3) the ability to analyze weld data across a nominal range of weld currents; (4) the development of a methodology capable of qualifying and quantifying the suitability of a training set; (5) the ability to accurately model weld acoustics despite the high degree of variability inherent in the weld data; (6) the detection and elimination of linear dependance between the input parameters; (7) increased mapping accuracy, greater stability, and faster convergence through the refinement of scaling procedures, selective use of activation functions, use of a modified learning algorithm, and the dynamic application of an appropriate momentum and learning coefficient strategy. Of particular interest was the informational characteristics of the weld acoustic data. Analysis indicated that the migration from a magnitude based weld acoustic model to a statistical model based on variability may be appropriate and merits investigation.

Key Words: Backpropagation; cluster analyses; information models; neural networks; optimization; rule extraction; sensors; weld acoustics

Introduction: During an arc welding process, arc instability is a common phenomena often associated with minor changes in voltage, current, base material, filler material, flux, or shield gas composition. These minor changes are not of themselves rejectable, or even detectable, but still may cause changes in metal transfer that can lead to rejectable weld defects. These problems are now detected, if at all, by the close monitoring of the weld by the weld operator. In fact, an experienced operator can maintain proper operating conditions by monitoring arc sound [Lancaster, 1987]. Additionally, it has been shown [Arata, et al., 1979] that the operator can discern droplet detachment events and arc stability acoustically. This capability allows the

experienced weld operator to closely control arc characteristics and identify potential defect producing events based on the sound of the weld. In automated robotics weld systems, this functionality is required from a minimally intrusive machine detection system that is acoustic based and sensitive to the same acoustic parameters as is the human ear. A non-linear, feed forward neural network trained to analyze weld acoustic data utilizing a backpropagation learning model has been successfully developed at the Carderock Division of the Naval Surface Warfare Center (CARDEROCKDIV/ NSWC), Code 2815 [Matteson, et al., 1992]. The neural network consists of three hidden layers and uses thirty average power spectra, one peak amplitude, and one RMS amplitude as parameters for a total of 32 input nodes (Figure 1). This trained neural network is the core of the Weld Acoustic Monitor (WAM) and is capable of discerning between acceptable or unacceptable weld conditions, in near real time, utilizing weld acoustic data [Matteson, et al., 1992]. The WAM is a sensor subsystem on the Programmable Automated Weld System (PAWS), a successfully demonstrated laboratory prototype, funded as part of an Advanced Technology Development project for the Navy, and currently being transitioned into shipyard use [Kline, 1992].
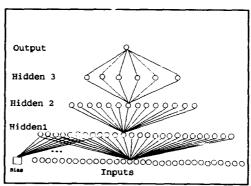
Figure 1. Initial WAM Design

Artificial Neural Networks (ANNs) are loosely based on current theories of how the human brain works, that is, through the interconnection of neuronal cells. A key advantage of ANNs over conventionally written software is the way in which ANNs imitate the brain's ability to make decisions and draw conclusions when presented with complex, noisy, irrelevant, and/or partial information. Another advantage is that ANN applications are not hand crafted programs, but rather the result of feeding training data to a network model, which then learns to output the desired results. Once a network has been successfully trained, it will ideally be capable of analyzing data that is different from that which it was originally exposed to during the training sessions. In other words, it is capable of generalization. This is a big advantage over conventional software which must be specifically programmed to handle every anticipated input in a sequential fashion.

Unfortunately, this body of knowledge is contained in an abstract set of neuronal cell connection weights which are highly sensitive to initial conditions defined before the learning process (initial ANN physical design, training model, and the choice of initial connection weights). Interpretation of these connection weights is complicated by the fact that neural networks are non-linear systems possessing a high degree of interdependence between nodes. This usually results in a "black box" approach to neural network design,

146

training methodology, and application development. The goal of this proposed study is the development of a methodology, that can be applied to this "black box", capable of extracting the underlying knowledge and associated rules, as expressed in the set of connection weights, in some more understandable representation. Of prime interest is the understanding of how networks achieve their mapping between the informational content of input data, weight assignments, and unit activations. This information could then be used in the optimization of the networks physical design, learning model, and initial connection weight estimations. Analysis and interpretation of neural network behavior is inherently difficult due to the high dimensionality of the solution space. Functional neural networks may consist of hundreds of nodes sharing thousands of connections.

A generic methodology, consisting of a series of non-parametric and para-metric statistical analyses is currently being developed at CARDEROCKDIV NSWC, Code 1253 under Independent Exploratory Development (IED) funding. For the purpose of this study a subset of this methodology, consisting of non-parametric analyses only, was conducted on both the WAM and the associated weld acoustic data. Weld acoustic data was analyzed for infor-mation content and characteristics. The WAM was then analyzed in the attempt to establish a direct mapping between hidden unit activations, weight clustering, and the information contained in the weld acoustic data. In summary, four aspects of WAM neural anatomy were analyzed: (1) informa-tional content and characteristics of the weld acoustic data; (2) hidden unit activations; (3) connection weights; (4) output activations.

Approach: Central to the idea of information modeling is the concept of clustering. Data presented to the WAM consists of a series of input vectors. If these input vectors were to be plotted in Euclidean space, they would form information clusters indicative of the states the neural network should be capable of distinguishing. This clustering is the result of similar vectors, as defined by the magnitude of their dot product, to be of small relative distance to each other. The Min-Max Theorem regarding neural networks states that the ability to increase the information content of the input data is a direct function of the ability to minimize the distance between vectors in an information cluster while maximizing the distance between information clusters. In other words, an optimum condition exists when individual information clusters are tightly focused but exist far apart from each other.

Preliminary attention was focused on the relationship between the informa-tional content and characteristics of the input data and the hidden unit activa-tions. The units that make up the hidden layer(s) can be thought of as "learned feature detectors" or "re-representation units" because the activity patterns in the hidden layer(s) are an encoding of what the network perceives as significant input features. Specifically, two questions were asked: "What is the quality and quantity of the information content of the input data?" and "What are the activation patterns of units in the hidden layer(s) in response to the information content of the input data?". A series of cluster analyses

147

consisting of Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), Canonical Discriminant Analysis (CDA), and Ward's Minimum Variance Cluster Analysis (WMVCA) were used to: (1) qualify and quantify information characteristics present in the weld acoustic data; and (2) define hidden unit activation patterns, in response to the input data, in the attempt to establish a direct mapping between these patterns and those of the data.

Difficulties immediately arise however, due to the high dimensionality of the problem space. Humans are intrinsically incapable of interpreting high dimensional spatial representations. A method of reducing the dimensionality of the problem space while minimizing the amount of information lost in the process is needed. Various cluster analyses applied simultaneously can reduce dimensionality and thus provide a useful set of analysis tools.

When HCA is applied to input data, it results in a tree of relational patterns. Similar patterns of information are closely related in a tree like structure while dissimilar patterns remain distant cousins. Distances between representational clusters are made roughly proportional to the distances these clusters maintain in hyperspace, resulting in a relative quantitative representation [Dennis and Phillips, 1991].

PCA is a statistical technique for calculating the major directions of variation of a set of data vectors in some high dimensional space where, as much as possible, the original distances between the vectors are preserved, in a least means sense [Dennis and Phillips, 1991]. It is used when no hypotheses have been formulated as to which dimensions constitute the most relevant information. When applied to input data, PCA will extract the dimensions along which the data vectors vary most, in the assumption that the directions of greatest variance will correspond to the most relevant information. In the cases where the major component of variation is noise, the analysis is rendered useless. PCA can also be applied to hidden unit activations during and after training. When applied during training, PCA demonstrates how the hidden unit activations cluster in response to training data. When conducted on a trained neural net, hidden unit activation clusters can then be mapped to those inherent in the input data.

CDA is another statistical technique used to compress high dimensional space into two or three dimensions so it can be easily visualized (Dennis and Phillips, 1991). In CDA each vector in the original space is designated as belonging to a group. This information is used to find the directions along which vectors within a group are clustered as tightly as possible while maximizing the between group separation. CDA is used to confirm or deny the hypothesis that the given groups are significant in the network's performance of the task. It should be noted that the following properties are inherent in CDA (Dennis and Phillips, 1991): projecting the original space onto the canonical variates not only rotates the original space but also distorts it. This is because the canonical variates are not constrained to be orthogonal, yet the canonical

148

variates are transposed to orthogonal axes. Thus two clusters may appear further apart than they were in original representational space.

As with PCA, CDA can be conducted on the input data as well as hidden unit activations, both during and after training, providing complimentary analysis techniques. PCA requires no *a priori* assumptions as to how the network performs the required task while CDA presupposes important categories. Should the network not use these categories then CDA will result in a distorted representation. When appropriate categories are chosen, however, CDA provides a much clearer representation as it generally results in clusters of higher definition. In cases where PCA results in the intersection of one or more information clusters, CDA provides an indication of the separability of the clusters in question.

Results: An Analysis of Variance (ANOVA) conducted on the WAM training set demonstrated that the data exhibited a lesser degree of variability between weld conditions than within. This condition was attributed to the high degree of variability inherent in the weld acoustic data. A new master training set was constructed from data, gathered during numerous weld sessions, conducted over a period of approximately a month, and at a nominal range of weld currents, that demonstrated an appropriate distribution of variability. It was noted that, despite the robust nature of the training set, minimal difference were observed between the measures of variability.

Cluster analyses were conducted on the master training set. HCA demonstrated an intricate relationship between individual weld acoustic data vectors representing the two weld conditions. Input vectors were grouped in a complex relationship of sub-classes where in any given sub-class, vectors representing both weld conditions were present. Similar results were obtained when PCA and CDA analyses were conducted. Sample results of these tests are contained in Figures 2 and 3. It can be observed that no well defined mapping exists in representational space. This was especially disturbing in the case of the CDA analysis which tends to distort the mapping causing clusters to appear further apart than they actually are. It should be noted that these graphical results are in keeping with those from the ANOVA which demonstrated negligible difference between variability within a given weld condition and variability between weld conditions. An Analysis of Covariance (ANCOVA) revealed that a significant component of

Figure 2. PCA - Nominal Range of Weld
Currents (0=Unacceptable, 1=Acceptable)

149
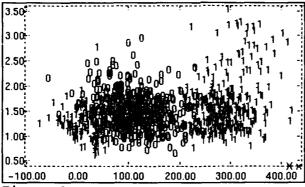
Figure 3. CDA – Nominal Range of Weld
Currents (0=Unacceptable, 1=Acceptable)
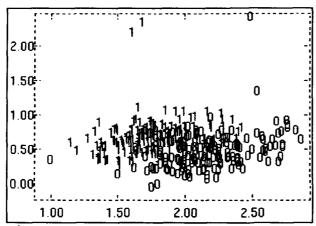
Figure 4. PCA – Medium Current Weld
(0=Unacceptable, 1=Acceptable)
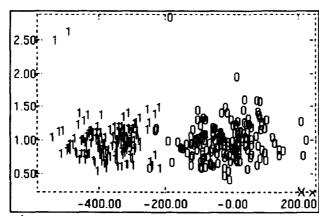
Figure 5. CDA – Medium Current Weld
(0=Unacceptable, 1=Acceptable)

the total variability was due to variation in weld current. It was reasoned that this component of variation could be controlled by the addition of a weld current input to the neuronal model.

In order to verify this observation, the master training set was decomposed into three subsets representing low (190 amps), medium (250 amps) and high (300 amps) weld currents. These sets were then subjected to nonparametric analysis. Sample results for medium current weld data can be seen in Figures 4 and 5. PCA analysis resulted in two relatively well defined, though intersecting, clusters. Results from CDA analysis indicated that though intersecting, these information clusters were separable. Medium weld current data exhibited the most pronounced cluster definition with minimal overlap while low and high weld current data exhibited clustering that was less defined and overlap that was considerably more pronounced. Discussions with various weld engineers indicated that this was in keeping with field observations. Unacceptable welds are noticeably more noisy and demonstrate a higher degree of variability than acceptable welds con-

150

ducted at the same weld currents. At high weld currents, the process is so noisy that it is difficult for humans to discern between weld conditions. Conversely, welds conducted at low weld currents produce low acoustic output, again making it difficult for humans to discern between weld conditions. Welds made at medium weld current were optimum for acoustically discerned weld conditions.
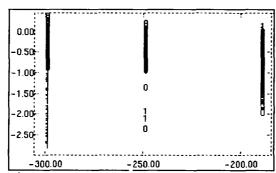
Figure 6. **PCA** - Nominal Range of Weld Current with Weld Current as Input Parameter (0=Unacceptable, 1=Acceptable)

Weld acoustic input vectors, gathered over a nominal range of weld currents, were modified to include their associated weld current values. Figure 6 depicts the results of a PCA conducted on this data set. It can be observed that this analysis indicated that the major directions of variation were present at 190, 250, and 300 representational units indicating that the major source of variability detected by PCA was due solely to variation in weld current, thereby validating the addition of a weld current input to the neuronal model. CDA yielded identical results. It should be noted that while representing a major improvement in the weld acoustic model, the addition of weld current as an input parameter did nothing to increase mapping definition within any given weld current class.

Investigations showed that no direct mapping between information clusters present in the weld acoustic data and hidden unit activations existed. This indicated that the major focus area for potential optimization remained in the data processing component of the WAM. After lengthy investigation, several modifications were made: (1) scaling factors related to dynamic range control were shown to decrease ANN mapping accuracy and were removed; (2) normalization schemes utilized were found to introduce linear dependence between input nodes and were modified to insure linear independence; (3) use of the normalized cumulative delta rule as a learning model resulted in minimized convergence time and an
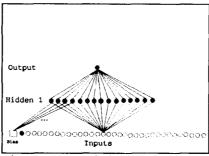
Figure 7. Optimized WAM Design
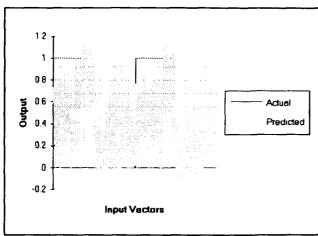
Figure 8. Original WAM Performance Characteristics

increase in mapping accuracy and generalization capability when used in conjunction with a dynamic application of an appropriate momentum and learning coefficient strategy.

The implementation of modifications indicated by the application of these non-parametric analyses resulted in the simplification of the neuronal model from three hidden layers to one, as depicted in Figure 7, with an associated reduction in processing time. Improvement in classification capability of the ANN is demonstrated by comparing and contrasting Figures 8 and 9.

Conclusion: Non-parametric analyses conducted on the weld acoustic neuronal model resulted in: (1) a simplification of the neural network from three hidden layers to one, with an associated reduction in processing time; (2) an increase in overall accuracy; (3) the ability to analyze weld data across a nominal range of weld currents; (4) the development of a methodology capable of qualifying and quantifying the suitability of a training set; (5) the ability to accurately model weld acoustics despite the high degree of variability inherent in the weld data; (6) the detection and elimination of linear dependance between the input
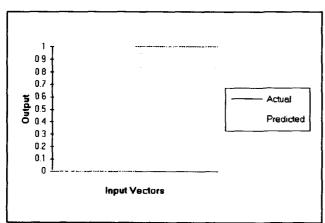
Figure 9. Optimized WAM Performance Characteristics

parameters; (7) increased mapping accuracy, greater stability, and faster convergence through the refinement of scaling procedures, selective use of activation functions, use of a modified learning algorithm, and the dynamic application of an appropriate momentum and learning coefficient strategy. In addition, these analyses indicated that: (1) optimization efforts should focus on input data selection and presentation; (2) informa-

152

tion content of the weld acoustic data is diminished by the high degree of variability inherent in a chaotic process such as welding. In light of this additional information, an in depth spectral analysis is being conducted on weld acoustic data. Initial results have indicated that a measure of this variability, appearing at frequencies above 10 K Hz, may provide an accurate and robust indication of weld condition.

While variability measurements appear to provide a useful indicator to discern between acceptable and unacceptable weld conditions, it is doubtful that such an indicator will prove useful in discerning between different types of unacceptable weld conditions such as porosity, lack of fusion or shield gas loss. Current research efforts are being focused on investigating the potential of a dual, neural network based approach to the analysis of weld acoustic data. In such a scenario, signal variance would be monitored for the detection of unacceptable weld conditions. If such conditions were detected, the signal would then be subjected to classification schemes that focus on spectral characteristics of the signal. An additional benefit of such an approach is that the qualification of signal variance used to discern between acceptable and unacceptable weld conditions would provide a relative measure of weld quality independent of magnitude. Fluctuations in signal strength, common in hostile industrial environments, would have no effect on the resultant analysis.

It should be noted that such an approach, based upon differences in magnitudes rather than the magnitudes themselves, would be in keeping with the biophysics of an actual neuron which is incapable of processing magnitude based data. Research has shown [Sejnowski and Lisberger, 1991.] that *"Neurons cannot represent absolute values of sensory data with high accuracy* because of their limited dynamic range-firing rates are typically from 1 to 100/second. Furthermore, statistical variability of the spike arrival time requires either time averaging or a population averaging to achieve even one significant figure of accuracy. *This limitation favors the representation of differences, rather than absolute levels."*

In addition, the effects of background noise on the current WAM model have never been accurately determined. If airborne acoustic, plateborne acoustic and arc current signals were monitored for fluctuations in variance, and compared, a criterion for discerning weld condition, independent of background noise, may exist. In summary, an optimized acoustic model is proposed that can potentially provide increased accuracy and generalization capability while being impervious to background noise and provide a relativistic measurement independent of signal level.

Bibliography: Arata, I. Y., Futamata, K., Toh, T., "Investigations on Welding Arc Sound, (Report II), Evaluation by Hearing Acuity and Some Characteristics of Arc Sound", Transactions of JWRL, Vol. 8, No. 2, 1979.

DeMenthon, D., Kamagar, F., Davis, L., Teolis, T., Matteson, A., "Learned Classification of Welding Modes by a Neural Network with Acoustic Inputs", University of Maryland, Center for Automation Research, August, 1989.

Dennis, S., and Phillips, S., "Analysis Tools for Neural Networks", Technical Report 207, Department of Computers, University of Queensland, Queensland, 4072, Australia, 1991.

Kline, M. D., "Programmable Automated Welding System (PAWS) System Overview", 3rd International Welding Research Conference, Gatlinberg, TN, June 1992.

Lancaster, J. F., "The Physics of Fusion Welding, Part 1: The Electric Arc in Welding", IEEE Proceedings, Vol 134, No. 5, September, 1987.

Matteson, M. A., Morris, R., Tate, R., "Real Time GMAW Quality Classification using an Artificial Neural Network with Airborne Acoustic Signals as Inputs", International Conference on Computerization of Welding Information, Orlando, FL, October, 1992.

Matteson, M. A., Raines, D., Morris, R., "An Optimal Artificial Neural Network for GMAW Arc Acoustic Classification", 3rd International Welding Research Conference, Gatlinburg, TN, June, 1992.

NeuralWare Inc, "Neural Computing", NeuralWare Inc, Pittsburgh, PA, 1991.

Sejnowski, T., J., and Lisberger, S., T., "Neural Systems for Eye Tracking", Naval Research Reviews, Vol XLIII, 1991.

Sokal, R., and Rohlf, F., "Biometry - The Principles and Practices of Statistics in Biological Research", Second Edition, W. H. Freeman and Company, San Francisco, CA, 1981.

Wasserman, P., "Neural Computing - Theory and Practice", Van Nostrand Reinhold, NY, NY, 1989.

Zeidenberg, M., "Neural Networks in Artificial Intelligence", Ellis Horwood Series in Artificial Intelligence, Ellis Horwood Ltd., 1990.

A SMART GENERIC SHOCK ABSORBER TEST STAND

Patrick J. Sincebaugh
U.S. Army Research Laboratory
Materials Directorate
Watertown, MA 02172-0001

Abstract: The imminent reductions in budget and personnel throughout the Department of Defense requires that new and innovative testing techniques be developed in order to test and maintain complex systems. The Army Research Laboratory Materials Directorate is currently developing a system that utilizes Artificial Intelligence technology to test shock absorbers for the M113 and Bradley Armored Vehicles. The shock absorbers will be dynamically tested utilizing a hydraulic test stand. The test stand provides a hardcopy of the following data: force, displacement, cycle time, and temperature. Currently, this data must be manually analyzed in order to evaluate the condition of the shock absorber. The scope of this project is to automate the testing process by utilizing a personal computer to acquire the data. The Smart Shock Absorber Test System will then utilize neural network technology to evaluate the condition of the shock absorber.

Introduction: Maintaining and testing complex systems is becoming a more challenging task due to budget and personnel cuts throughout the Department of Defense. In order to meet this challenge, new and innovative testing techniques must be developed. The U.S. Army Research Laboratory Materials Directorate has been doing research in developing Smart Systems to help meet testing needs throughout the Army. A Smart System can be defined as a computer based system that utilizes state-of-the-art technology, often Artificial Intelligence(AI) technology, to enable the system to make decisions and/or perform functions that were previously made by human operators. The development and implementation of Smart Systems have shown high Returns-On-Investment throughout a wide variety of industries, including airlines, aerospace, banking, and government agencies[1]. The implementation of Smart Systems throughout the Department of Defense can help ease the problems associated with budget and personnel reductions.

155

The U.S. Army is currently developing a Smart System that will be used to test shock absorbers for the M113 and Bradley Armored Vehicles. The shock absorbers will be dynamically tested using a hydraulic testing device. The hydraulic device will oscillate the shock absorber at 100 cycles per minute, over a 3 inch stroke. At every 0.0219 seconds(for this test case) the device provides the following data: displacement, force, time, temperature, and velocity. This data is acquired by a personal computer via an analog to digital(A/D) converter board that plugs into a PC's expansion slot. The data will then be analyzed by a software package that will be developed utilizing neural network technology. The Smart System will provide a diagnostic output that classifies the shock absorber as either nominal or faulted. The system will also provide an interactive interface that will allow the operator to examine the relationship between the force and displacement data for the shock absorber. Another feature of the Smart Shock Absorber Test Stand is that it will be adaptable to test different types of shock absorbers without software and/or system modifications. The development of "generic" type systems is a necessary approach in order to reduce future system development costs in an era of reduced budgets.

Problem Statement: The need to develop an improved testing methodology for the M113 and Bradley shock absorbers was identified at the Red River Army Depot(RRAD). The Remanufacturing Facility at RRAD is responsible for the maintenance and overhauling of Armored Fighting Vehicles(AFV). When the vehicles arrive for maintenance, they are completely disassembled. The individual system components, such as the track, engine block, shock absorbers, etc., are then tested. The vehicle is then reassembled with components that have passed the respective individual tests. The reassembled vehicle is then performance tested by driving it around a test track for a specified amount of time and distance. If the vehicle meets or exceeds all of the performance criteria, it is released back into the field. During the disassembly process, there are numerous functional tests that are performed on the individual system components. However, used shock absorbers are reinstalled or discarded after a visual inspection, without the benefit of a functional test. The Army Audit Agency has confirmed high field failure rates for the shock absorbers, and has consequently recommended that a diagnostic test for the shock absorbers be developed.

A hydraulic testing device was procured in order to provide functional testing capabilities for the M113 and Bradley shock absorbers. The test stand consists of a test console

and a hydraulic power supply, which supplies hydraulic fluid to a servo cylinder mounted on a load frame. A shock absorber is mounted vertically into the load frame, and is subjected to a sinusoidal motion of 100 cycles per minute(adjustable from 0-290 cycles/minute) at a 3 inch stroke. The test stand provides the following shock absorber data: resistance(force), temperature, extension(displacement), and the cycle rate. Currently, this data is provided on a paper printout. The operator must manually analyze the data from the hardcopy. The time it takes to make a pass/fail decision utilizing this technique is not conducive to a fast paced production line environment.

Proposed Problem Solution:
Automating Data Acquisition and Analysis Process:

The first step in transforming the hydraulic test stand into a Smart Shock Absorber Test Stand is to automate the data analysis process. The original test stand sends the data to a printer, where a hardcopy is provided. The data signals that are sent to the printer are analog signals. In order for a computer to be able to acquire and analyze this data, it must be converted to digital format. This is done by rerouting the signal lines into a Tecmar analog to digital(A/D) converter board which plugs into an expansion slot of an IBM PC. The Tecmar board allows for 16 single ended or 8 true differential channels of analog to digital conversion with 12 bit resolution. This data transfer process is controlled by a software package that was written in-house using the C programming language.

Data Analysis: Once the data is in digital form, it can be manipulated and analyzed by the computer and its associated software packages. The data that will be output from the test stand to the computer is the shock absorber displacement, the force on the shock absorber, and the cycle rate. The relationship between the shock absorber displacement and the force on the shock absorber is often analyzed by the shock absorber Original Equipment Manufacturer(OEM) to determine if the shock is nominal or faulted. Because of the relevance of this force and displacement data, a software function XY_PLOT has been written that plots the Force vs. Displacement data on an xy scale. The function XY_PLOT is not limited to plotting just force and displacement data, it is robust enough to accept any data in the following format:

```
7.35    3215
6.97    3456
2.45    1254
 .       .
 .       .
 .       .
```

where the first column represents the x-axis data, and the second column represents the y-axis data. The robustness of the function XY_PLOT is important because the test operator may be interested in plotting other data parameters. Other data parameters of interest are shown in Figures 1,2 and 3.

The revised shock absorber test system will plot the force vs. displacement data after 143 data points have been gathered. A sorting routine analyzes the data, and calculates the appropriate x-axis and y-axis ranges. Future work may consist of plotting the data points in real-time, but currently the advantages don't outweigh the disadvantages of this approach.

Figure 4 shows a plot of the shock absorber force vs. the corresponding shock absorber displacement for both a nominal and a faulted shock. A nominal shock absorber should result in a force vs. displacement plot that is symmetrical, and somewhat elliptical. Deviations in the force vs. displacement plot can be analyzed to pinpoint particular problems with the shock absorber. For example, a deviation in the plot when the shock is being stroked towards the extended position may indicate excessive fluid loss. A deviation in the plot when the shock is being stroked in the collapsed position may be the result of physical damage, often caused by stones and other debris.

The OEM also recommends examining the force parameter for the shock absorber during dynamic testing. Each shock absorber has a new-part tolerance band for both compression and extension forces. Forces that fall outside of the specified range may be the result of damage and/or wear associated with the shock absorber. This condition will be monitored via the software function FORCE_TOLERANCE. This function will allow the operator to interactively change the tolerance limits if the default values are not satisfactory. This will allow for different types of shock absorbers to be tested without making system software modifications. The function FORCE_TOLERANCE will read the force data that is acquired from the test stand and compare the value with the specified tolerance range. Statistics will be kept as to how many data values are in/out of the specified range. After the test is complete, the operator can view these statistics and decide whether to accept/reject the shock absorber, or whether the force tolerance band needs to be changed.

Development of a Smart System: The revised shock absorber test system will not only automate the data acquisition and analysis process, but will also recommend a pass/fail decision to the operator. The smart system will be developed utilizing neural network technology. Neural

networks are being used because of their inherent parallel computation capabilities, which is a necessary characteristic for any methodology that must be performed in a timely fashion. Another positive aspect of utilizing neural networks is their ability to solve problems associated with classification and/or pattern recognition. Since the testing methodology for an Armored Vehicles shock absorbers consists of analyzing the patterns associated with force vs. displacement data, the problem lends itself directly to a neural network solution. The use of neural networks also allows the system to be flexible enough to be adapted to test different types of shock absorbers by providing an interactive interface in which the operator can retrain the neural network. By using this approach, no software revisions are necessary, thus the user avoids having to recompile the software. This alleviates the problem of requiring a computer programmer to be present in order to modify the test system.

The intelligence of the system will be implemented by developing a backpropagation neural network. The backpropagation model has become the most widely accepted model over the past 5 years. A neural network customer survey has revealed that approximately 60% of the neural network applications developed utilize the backpropagation algorithm[2]. An example of a simple backpropagation network is shown in figure 5. The backpropagation model consists of atleast three layers: an input layer, a hidden layer(sometimes more than one), and an output layer. Each layer consists of multiple processing elements(PE's), also known as nodes. A node is analogous to the biological neuron in the brain. Each PE has multiple input paths(analogous to dendrites for you biological types), each having a weight value associated with it. An individual node is shown in figure 6. Each node has an internal activation that is calculated from the input values and weights using the following formula:

$$Y = \Sigma X_i W_i$$

The calculated value(Y) is then modified by the transfer function(F(Y)). Many different transfer functions can be implemented as long as they are differentiable and monotonically increasing. The authors have achieved the best results using the hyperbolic tangent function. The output of the node is then either input into the next layer(for an input node), or is output as the networks response to the corresponding input value.

The backpropagation network for the shock absorber test system learns how to classify faulted and nominal shocks through a supervised learning process. When the backprop network begins training, the initial weight values of each

159

processing element are assigned a random value. In supervised learning, the network is presented an input value in which the corresponding output is known. The networks output value is then compared with the desired output value. If the Root Mean Square(RMS) of the error is within a specified range, then the network is considered to be trained. Otherwise, the weights of the individual nodes are adjusted to decrease the error. This learning algorithm is known as the Delta Rule. The network will go through an iterative process until the RMS error converges to the specified level.

The most important step in developing a smart system is to first understand the relationship between the data and the systems output. For the shock absorber test system, this relationship was examined by conducting interviews with shock absorber experts, and by graphically analyzing sample data obtained by the hydraulic test stand. It was determined that the force and displacement data would be used to train a neural network to classify faulted and nominal shock absorbers. Once the relevance of the data is characterized, it must be massaged and manipulated before it can be used to train a neural network. The force and displacement data is first scaled using the software function Data_scale. The input data is scaled to within the range of -0.6 to 0.6. In general, if more input values are used, then the data should be scaled to within a smaller range. A future enhancement of this system will take into account the number of inputs, and scale the data accordingly.

Once the input has been massaged, it is input into the neural network. Preliminary studies have used a neural network with 246 input nodes(143 data points for both the force and displacement values), 18 hidden nodes, and 2 output nodes. The output nodes correspond to either a faulted or nominal shock absorber.

Once the network is designed it must be trained with data in which the desired output is known for each input value. For the shock absorber problem, it is important to train the network with shock absorber data that represents all of the types of conditions that the system should recognize. The training data for the neural network(only 4 input values for this example) must be in the following format:

    4.56 7.56 4536 3457 1.0.0.0
    2.34 5.43 3476 3768 0.0 1.0
      .   .   .   .   .   .
      .   .   .   .   .   .
      .   .   .   .   .   .

The first two columns represent displacement input values. The next two columns represent force input values. The

last two columns represent the desired output for the corresponding input values. The training file can be developed by calling the software function FILE_FORMAT after the 'Train Neural Network' option is selected from the main menu. A shock absorber with known characteristics must be setup in the hydraulic test stand. The output columns must then be appended by the user using an ASCII text editor.

The next step after training is complete is to test the network by using input for which the outcome is known, but not presented. The file format for the testing data(for a network with only four inputs) is as follows:

```
6.76 7.56 3034 4532
5.23 5.34 5472 2354
   .     .     .     .
   .     .     .     .
   .     .     .     .
```

Each column of numbers is representative of an input value. The first two columns in this example represent displacement values, while the second two columns represent force values. The actual file for the shock network would contain 246 columns of data. Once the test file is formed, it should be used as input to the neural network. The networks output should then be compared to the desired output.

If the output is satisfactory, then the neural network will be converted into C code. The resulting function is then integrated with the system software package. If the output is not satisfactory, then the neural network must be reexamined. Many times the inconsistencies of the neural network are the result of poor data representation. The problem could be due to data that is not representative of the overall problem. Another problem area could be with the preprocessing of the data. There are many good papers that deal with this problem[4,5,6]. There are also many network parameters that can be manipulated to obtain better results(number of hidden nodes, learning rates, etc.). The neural network manual should be consulted for further information.

Conclusions: Innovative testing techniques are being developed within the Army in order to cope with the problems associated with budget and personnel reductions. The authors have described a Smart Shock Absorber Test Stand that will be used to test M113 and Bradley shock absorbers. The system utilizes neural network technology to analyze force and displacement data that is obtained from a hydraulic test stand. After the data is analyzed, the system will output a diagnostic decision regarding the

condition of the shock absorber. By implementing a functional testing methodology for the M113 and Bradley shock absorbers, the Army should reduce in-field failure rates, thus increasing combat readiness while decreasing maintenance costs.

References:

[1] Graettinger, T., Hitt, B., Klimasauskas, C., _Applying Neural Networks in Business, Industry and Government_, NeuralWare Inc., Pittsburgh, PA, 1989

[2] Caudill, Maureen, "The View From NOW", _AIExpert_, June 1992, pp. 24-31

[3] Crooks, Ted, "Care and Feeding of Neural Networks", _AIExpert_, July 1992, pp. 36-41

[4] Stein, Roger, "Selecting Data for Neural Networks",_AIExpert_, February 1993, pp. 42-47
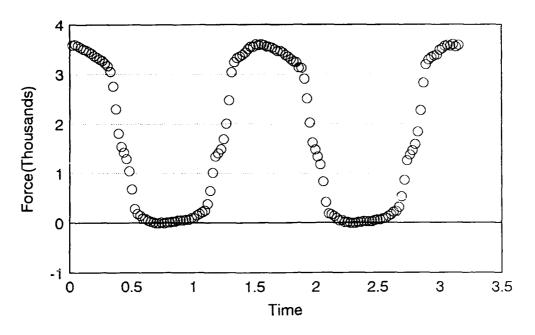
Figure 1. Force vs. Time Plot of Shock Absorber Data
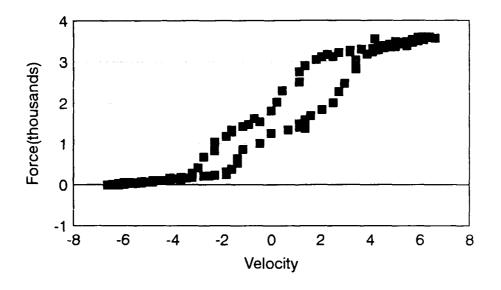
Figure 2. Force vs. Velocity Plot of Shock Absorber Data

163
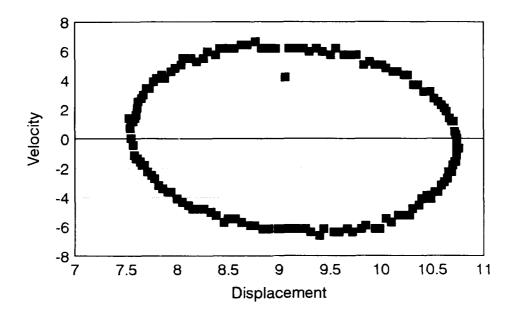
Figure 3. Velocity vs. Displacement Plot of Shock Absorber Data

■ Faulted shock Absorber
⬚ Nominal Shock Absorber
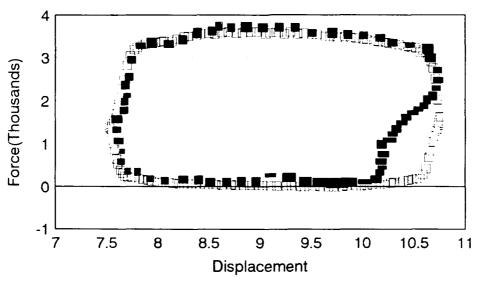
Figure 4. Force vs. Displacement Plot of Shock Absorber Data

164

RIPPLE-LOAD CRACKING IN THREE ALLOY SYSTEMS

Robert A. Bayles, Peter S. Pao, Steven J. Gill
Code 6320
Naval Research Laboratory
Washington DC 20375

George R. Yoder
Code 1131
Office of Naval Research
Arlington VA 22217

Abstract: Designs which assume no crack growth below K_{ISCC} may not be conservative if a small cyclic "ripple" load is experienced by the structure but arbitrarily disregarded by the designer or operator. Experiments are described in which a ripple-load effect was observed in a steel, a titanium alloy, and an aluminum alloy. A method is described which predicts the results of long-term ripple-load experiments using data from relatively quick corrosion-fatigue experiments. Evidence to date indicates that materials which are relatively more resistant to stress corrosion are more vulnerable to the ripple-load effect.

Key Words: ripple-load; stress-corrosion; corrosion-fatigue

Introduction: Stress-corrosion cracking (SCC) is a mode of subcritical crack growth which will occur if a sensitive material is exposed to a corrosive environment under sufficient stress for a sufficient length of time. For structural materials which contain a crack (or crack-like defect), resistance to SCC is normally expressed in terms of the fracture mechanics parameter, K_{ISCC}, the threshold stress-intensity factor below which crack extension will not occur. Designs of structures based on K_{ISCC} as a parameter below which no crack growth will occur assume sustained or constant load conditions, or that any superimposed load fluctuations are insignificant. Although small fluctuations might seem insignificant, preliminary study has shown that their effect, called the "ripple effect" by Speidel [1], can be sizable. Recent work on steels, titanium, and aluminum alloys has suggested that the presence of such ripple loads can reduce the threshold for cracking substantially below K_{ISCC} and can shorten the life of a structure [2-5]. Parkins [6,7] has observed that small fluctuating loads _may_ produce SCC at significantly lower stresses than those required to produce SCC under purely static loads. Not all materials, however, appear susceptible to ripple-load degradation. For instance, Crooker et al [2] demonstrated, for a ripple load of 10 per cent of the maximum, a 60 per cent degradation from the static K_{ISCC} level in the case of a 5Ni-Cr-Mo steel, yet absolutely no degradation in the case of a 4340 steel.

This paper describes the initial experimental work on steels, in which the ripple-load effect was observed in "direct experiments". The development of a straightforward predictive methodology, based on corrosion-fatigue, and its application to a titanium alloy and an aluminum alloy is then presented. A design parameter, K_{IRLC}, which reflects a material's behavior under ripple-load conditions is defined.

Materials: An SCC-resistant alloy and an SCC-susceptible alloy from each of the three major families of structural alloys were included in this study. For the ferrous family, 5Ni-Cr-Mo-V and AISI 4340 steels were selected to represent, respectively, the SCC-resistant and SCC-susceptible materials. Similarly for the titanium family, a beta-annealed (BA) and a recrystallization-annealed (RA) Ti-6Al-4V, were selected as relatively SCC-resistant and SCC-susceptible microstructures. And for the aluminum family, the SCC-susceptible, peak-aged Al 7075-T651, which is well known for its low K_{ISCC}, and the SCC-resistant overaged Al 7075-T7351 were selected for study, both in the short-transverse (ST) orientation. Specific chemical analysis, product form, and mechanical properties for these alloys are available from previous publications [3-5].

Results and Discussion: In this paper, ripple-load cracking is treated as a high stress ratio, corrosion-fatigue phenomenon. The critical conditions and the predictive methodology for ripple-load effects involve the interface between SCC and corrosion-fatigue behavior. Parameters associated with SCC and corrosion-fatigue, such as K_{ISCC}, K_{max}, ΔK_{th}, ΔK, and R are used throughout the analysis.

(A) Ripple-load Cracking in Steels: Figure 1 shows the design of the "direct experiment" used to evaluate the effect of ripple loading on steels. The apparatus is a cantilever bend load frame, modified with a motorized cam to superimpose a small oscillating load onto the dead-weight load.

The specimens were fatigue-precracked in air, then the environment cup was sealed to the specimen. After the specimen was mounted in the load frame the cup was filled with 3.5% salt water and zinc anodes were coupled to the specimen. After 24 hours the dead-weight load was gradually applied while a crack mouth opening gage was used to determine, by means of compliance, the crack depth. Enough load was then applied to produce the desired K_{max}. The eccentric cam and spring apparatus was set up to cyclically reduce the load by the desired amount - in the case described here 10% of the dead-weight load. The cam motor was switched on to begin the experiment. A cyclic frequency of 0.1 Hz was used to simulate ocean wave motion. Evaporation losses were made up with distilled water as needed and the saltwater was replaced weekly.

Figure 2 shows the effect of ripple loading on SCC-resistant 5Ni-Cr-Mo-V steel. The predicted ripple-loading time-to-failure curve obtained, as described later in Section B, through integration of corrosion-fatigue data for 5Ni-Cr-Mo-V steel [9] is included in Fig. 2. As can be seen from Fig. 2, 5Ni-Cr-Mo-V steel, though resistant to SCC, is very susceptible to ripple-load cracking under a 10% ripple. The predicted ripple-load cracking threshold, K_{IRLC}, is only 31 MPa$\sqrt{m}$. This is much lower than the static K_{ISCC} of 110 MPa$\sqrt{m}$. This opens a large window for ripple-load cracking susceptibility, with a maximum potential degradation of 72%. The predicted time-to-failure curve under ripple loading agrees well with the experimental data.

Each of the open circles indicates a separate direct experiment. The longest direct experiment was run for 8000 hours and failure of the specimen was not observed. The prediction indicates that the true threshold value for ripple loading, K_{IRLC}, is substantially lower, but a direct experiment duration of much more than 30,000 hours would be required to confirm this.

168

For SCC-prone, high strength AISI 4340 steel, the K_{IRLC} predicted from direct experiments and from integration of the corrosion-fatigue curve, as described in Section B, is 33 MPa$\sqrt{m}$, which is identical to the K_{ISCC}. Therefore the susceptibility window is nonexistent and no ripple-load effect is expected.

(B) Analysis of the Ripple-Load Effect: A structure stressed above K_{ISCC} and under a sustained load is expected to fail by a stress-corrosion cracking mechanism. The addition of small ripples may accelerate the cracking process and shorten the anticipated useful life. A superposition model has been successfully developed to address the combined influence of cyclic and sustained loads in the regime above K_{ISCC} [8].

In this study, our attention was focused on the regime below K_{ISCC} where propagation of existing cracks and failure are not expected under a constant load condition. Thus, with the presence of small ripples superimposed on a large sustained load, the maximum stress intensity in the ripple-load cycle was equal to or less than K_{ISCC}. That is, the first condition for ripple-load cracking can be set as:

$$K_{max}^{RL} \leq K_{ISCC} \qquad (1)$$

Next, from corrosion-fatigue considerations, crack propagation is not going to take place during ripple loading unless ΔK^{RL} in the ripple cycle equals or exceeds ΔK_{th}:

$$\Delta K_{th} \leq \Delta K^{RL} \qquad (2)$$

$$\text{or} \qquad \frac{\Delta K_{th}}{1-R} \leq K_{max}^{RL} \qquad (2a)$$

Thus, a new parameter, K_{IRLC}, the ripple-load cracking threshold below which ripple-load cracking does not occur, can be defined as:

$$K_{IRLC} = \frac{\Delta K_{th}}{1-R} \qquad (3)$$

Combining (1), (2a) and (3), the conditions for a material to exhibit ripple-load cracking are:

$$K_{IRLC} \leq K_{max}^{RL} \leq K_{ISCC} \qquad (4)$$

Relation (4) is illustrated in Figures 2 and 4-7 for the various alloy systems. The region whose upper bound is the stress-corrosion cracking threshold, K_{ISCC}, and whose lower bound is the ripple-load cracking threshold, K_{IRLC}, defines a "window of susceptibility" in which the ripple-load effect would be anticipated. The wider the window, the more susceptible the material is to ripple-load cracking. In the extreme case, where K_{IRLC} approaches K_{ISCC}, the susceptibility window does not exist and no ripple-load effect is expected.

169

If one considers the difference between the threshold for ripple-load cracking K_{IRLC} and K_{ISCC}, then the maximum extent of ripple-load degradation can be defined as:

$$\% \text{ degradation} = (1 - K_{IRLC} / K_{ISCC}) * 100 \qquad (5)$$

Finally, the ripple-load cracking time-to-failure curve can be obtained from a simple piece wise numerical integration of the corrosion-fatigue crack growth rate curve [3], for the particular structural geometry of concern. For the 5Ni-Cr-Mo-V steel, the time-to-failure curve was predicted from cantilever bend bar geometry, and for the other materials, from a compact tension geometry.

To generate the corrosion-fatigue curve while simulating a ripple-load condition, and to measure threshold levels of stress-intensity range directly, precracked specimens were cyclically loaded at room temperature in a 3.5% NaCl solution with a stress ratio (minimum: maximum) of R = 0.90 (10% ripple loading), a haversine or a triangular wave form, and a cyclic frequency of either 0.1 or 5 Hz. Fig. 3 shows schematically the corrosion-fatigue apparatus. SCC thresholds were determined in the 3.5% NaCl solution using either constant load cantilever bend bar tests or slow-strain rate tests with a loading rate of 10^{-4} MPa$\sqrt{m}$/s. Crack lengths were determined using a compliance related CMOD technique.

The method described here uses corrosion-fatigue data from one specimen, typically obtained in a few weeks, to predict the complete ripple-load curve between K_{ISCC} and K_{IRLC}. The direct experiment approach required many experiments and, for steel, test durations of up to many years.

(C) Ripple-Load Cracking in Titanium Alloys: The predicted ripple-load cracking curves for two titanium alloys which exhibit different levels of SCC resistance are shown in Figs. 4 and 5. The SCC-resistant, beta-annealed Ti-6Al-4V has an SCC threshold of 60 MPa$\sqrt{m}$. The predicted K_{IRLC} is only about 28 MPa$\sqrt{m}$. As shown in Fig. 4, a large susceptibility window, representing a 53% ripple-load degradation, exists for this SCC-resistant titanium alloy. Experimental data illustrate a good agreement with the predicted time-to-failure curve.

Figure 5 shows the ripple-load degradation of the recrystallization-annealed Ti-6Al-4V which is less SCC resistant than the beta-annealed Ti-6Al-4V. The ripple-load cracking threshold was determined to be around 39 MPa$\sqrt{m}$, which is about 9% lower than the static K_{ISCC}.

(D) Ripple-Load Cracking in Aluminum Alloys: The predicted ripple-load cracking curves of SCC-resistant and SCC-prone aluminum alloys are presented in Figs. 6 and 7, respectively. Figure 6 shows the ripple-load time-to-failure curve of overaged 7075-T7351, which exhibits excellent SCC resistance, even in the short-transverse (ST) orientation. Like SCC-resistant ferrous and titanium alloys, SCC-resistant overaged 7075-T7351 has a large susceptibility window. The predicted K_{IRLC} is 58% lower than K_{ISCC}.

The ST-oriented, peak-aged 7075-T651 is well known for its low SCC resistance and has accounted for the bulk of SCC failures in high strength aluminum alloys. Yet, like the SCC-susceptible AISI 4340 steel, this peak-aged 7075 does

170

not exhibit any ripple-load degradation (Fig. 7). The K_{IRLC} and K_{ISCC} are identical in ST-oriented 7075-T651.

Summary: The ripple-load cracking susceptibility and the extent of ripple-load degradation of the six alloys studied are summarized in Table I. Table I clearly demonstrates that those materials which exhibit greater SCC resistance under static load conditions are far more susceptible to ripple-load degradation. Without exception, the SCC-resistant materials, ranging from 5Ni-Cr-Mo-V steel to beta-annealed Ti-6Al-4V to overaged 7075-T7351, are more prone to ripple-load degradation than the less SCC resistant materials. The significance of this finding is obvious, at least phenomenologically, as a material selected for its superior SCC resistance may fail if ripple-load conditions exist. To circumvent this problem, it is suggested that the ripple-load cracking threshold, K_{IRLC}, should be considered along with the static K_{ISCC} to determine allowable stress and inspection intervals.

Table I. Summary of SCC and RLC Properties

Material	SCC Resistance	RLC Susceptibility	RL Degradation
5Ni-Cr-Mo-V	High	High	72 %
AISI 4340	Low	None	-
BA Ti-6Al-4V	High	High	53 %
RA Ti-6Al-4V	Moderate	Low	9 %
(ST)7075-T7351	High	High	58 %
(ST)7075-T651	Low	None	-

A few words are also in order regarding the ripple-load time-to-failure curves. It is significant to note that the predicted ripple-load time-to-failure curves not only agree well with the experimental data but also permit the saving of the much greater time and expense associated with the direct experimental determination of such time-to-failure curves. For instance, as shown in Fig. 2, a test duration up to 30,000 hours (~ 3.5 yr.) would be required to establish experimentally the ripple-load time-to-failure curve for 5Ni-Cr-Mo-V steel.

Ripple-load cracking characteristics can be affected by mechanical and environmental parameters. However, many of these are not adequately understood. The size of the ripple can significantly influence the ripple-load cracking phenomenon [3]. Extremely small ripple loads (less than 2.5% of the sustained load) were found to have no damaging effect on 5Ni-Cr-Mo-V steel. Temperature and ripple-load frequency, which are known to affect corrosion-fatigue crack growth kinetics, should influence ripple-load cracking.

171

Conclusions:

1. Ripple-load cracking can be approached successfully as an extreme case of corrosion-fatigue. Critical conditions for ripple-load cracking have been defined.

2. Ripple-loading can significantly reduce the threshold for failure for SCC-resistant alloys while having little or no damaging effect on less SCC-prone alloys.

3. A new parameter, K_{IRLC}, the threshold stress intensity factor below which ripple-load cracking will not occur, is identified and recommended as a design consideration if ripple-load conditions are suspected.

4. The "window" for ripple-load cracking susceptibility is bounded at the top by K_{ISCC} and at the bottom by K_{IRLC}. Alloys more susceptible to ripple-load cracking will exhibit larger windows.

5. Ripple load time-to-failure curves can be predicted by a simple piecewise numerical integration of corrosion-fatigue curves. The predicted curves agree well with the direct measurements and afford significant time and cost savings.

References:

1. M.O. Speidel, Corrosion Fatigue of Steam Turbine Blade Materials, (New York, NY: Pergamon, 1983), p. 1.

2. T.W. Crooker, J.A. Hauser II and R.A. Bayles, Environmental Degradation of Engineering Materials, (University Park, PA: Pennsylvania State University Press, 1987), p. 521.

3. P.S. Pao, R.A. Bayles and G.R. Yoder, Trans. ASME, J. Eng. Matls. Tech., Series H, 113, 125 (1991).

4. G.R. Yoder, P.S. Pao, and R.A. Bayles, Scripta Met., 24, 2285 (1990).

5. P.S. Pao, S.J Gill, R.A. Bayles, and G.R. Yoder, Scripta Met., 24, 2085(1991).

6. R.N. Parkins, B.S. Greenwell, Metal Sci., 8/9 1977: p. 405.

7. R.N. Parkins, P.M. Singh, Corrosion, 46, 485(1990).

8. R.P. Wei, J.D. Landes, Materials Research and Standards, 9, 25(1969).

9. O. Vosikovsky, J. Testing and Evaluation, 6,175 (1978).
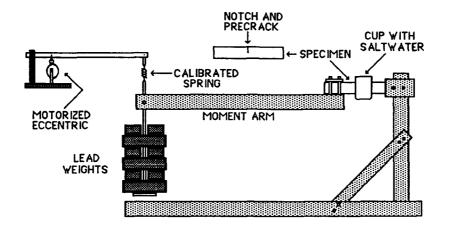
Figure 1. Ripple-load "direct experiment"
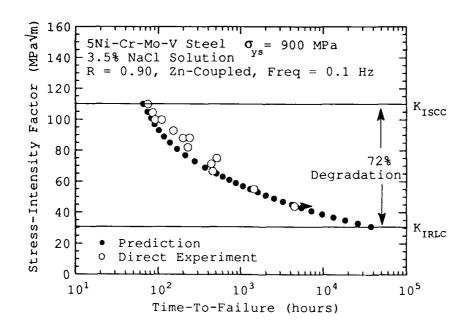
Figure 2. Ripple-load degradation in 5Ni-Cr-Mo-V steel
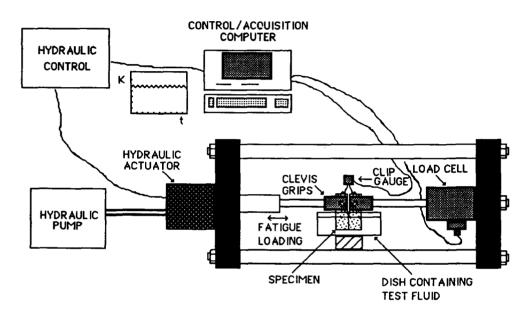
Figure 3. Corrosion-fatigue test machine
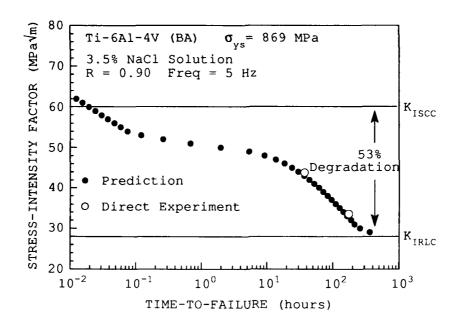
Figure 4. Ripple-load degradation in beta-annealed Ti-6Al-4V
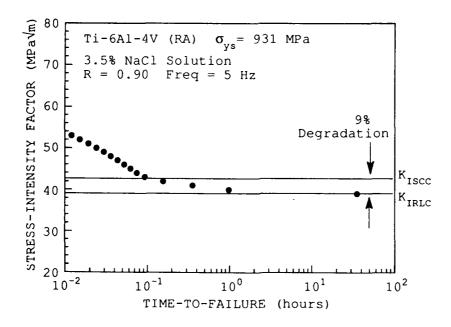
Figure 5. Ripple-load degradation in recrystallization-annealed Ti-6Al-4V
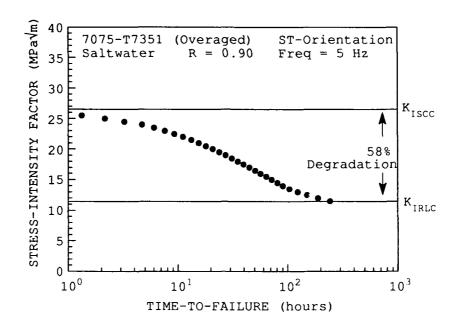
Figure 6. Ripple-load degradation in ST-oriented, overaged 7075-T7351
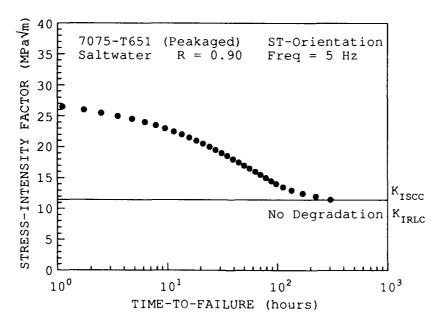
Figure 7. Ripple-load effect in ST-oriented, peak-aged 7075-T651

176

Wear Debris Characterization Combined with Mathematical Pattern Recognition Techniques for Condition Monitoring of Tribological Systems

K. Wolf and J. V. Czarnecki
Wehrwissenschaftliches Institut fuer Materialuntersuchungen (WIM)
Landshuter Str. 70,
8058 Erding, Germany

ABSTRACT

For condition monitoring and early failure detection of tribological systems with highly stressed components, like modern jet engines, it is found, that magnetic plugs in combination with wear particle characterization by SEM/EDX analysis are of increasing importance. Particle characterization can be used, because wear characteristics and particle features are related. The particle composition gives the information about a possible particle source. Particle morphology and size depend on the wear mode (fatigue, cutting, pitting, abrasion, adhesion, tribo corrosion). It is the objective of this paper to demonstrate how multivariate statistics and mathematical pattern recognition techniques (Principal Component Analysis and Hierarchical Cluster Analysis) applied to SEM/EDX results can translate the element composition of wear particles and other available wear related information into the identification of the alloy of a failing part and its localization in the tribosystem. The statistical interpretation of the data allows a failure pattern recognition. First results are reported.

INTRODUCTION

Increasing complexity of aircraft engines demands for improvements with regard to safety of operation. Early detection of problems and a fast availability of results of investigations can maximize operational safety and reduce repair costs. Therefore there is a necessity for diagnostic techniques that can monitor critical engine parts. A well established practice for condition monitoring concerning a precautionary measurement with respect to safety, reliability and operating life of jet engines is the detection of wear particles by magnetic plugs and subsequent "Debris Test" [1,2]. It is a mature and proven technique to quantify the amount of debris collected. Additional analysis of the particles by scanning electron microscopy (SEM/EDX) helps to distinguish wear in elemental composition, size and morphology. Wear can be caused by adhesive, abrasive, spalling and tribo corrosion processes. However, for the appearance of a mechanical problem there are many possibilities. Highly stressed parts commonly made of alloyed steel e.g. bearings are critical components with profound effects on the efficiency of a mechanical system and often fail due to fatigue. Metallic wear particles produced if lubrication becomes insufficient are consequently the most common wear present in the oil. Processes like that not only can lead into failure directly, but

177

mostly are precursors for more severe damages such as spalling, cracking, etc. Because of the various possibilities in a coplex system it is not easy to determine the location and the part where a damage starts. This last reason is important for condition monitoring investigation without a detailed knowledge of the tribo system. This paper tries to show how an approach combining the classical techniques (magnetic plugs, debris test, SEM/EDX-analysis) with mathematical pattern recognition techniques can help to improve condition monitoring and can lead to an expert system. The investigation bases on a 3-shaft jet engine which was intercepted by condition monitoring and the objective to locate the damage.

RESULTS

Principles of operation and material

The jet engine monitored concerning wear is modern, modular designed and equipped with magnetic plugs for condition monitoring in all modules (Figure 1). The tribosystem is lubricated by pressure circulation. One oil pressure pump supplies the bearings and the scavenge oil pump deliveres the oil to the oil tank. The filter boxes are provided with magnetic plugs and filters. Table 1 and 2 show the composition of alloys used for oil lubricated components, which produce wear in this mechanical system. Based on the Debris Test results the magnetic plugs of two out of 16 modules were investigated in detail by SEM/EDX. The plugs are designated as black and orange (Figure 1) and monitor the external gearbox and the rear bearing chamber including bearings and sealings. Table 3 shows the "Curriculum" of the jet engine or rather its modules referring to Debris Test results. The amount of magnetic particles captured by the magnetic plugs indicated the beginning of a damage. An early failure was first observed after 109.55 hours of operation. After an additionally test run (110.25h; 30min) increasing metallic wear was produced, confirming the first warning.

Fractography

Figure 2 shows the procedure how metallic wear particles were prepared for SEM/EDX-investigation. For Debris Test, wear particles are transferred from the magnetic plug to an adhesive tape by pressing the plug into the tape surface. The wear particles are removed from the adhesive with a solvent. After cleaning, the wear particles are cought with a magnet on a carbon target. This procedure guarantees that a representative selection of wear particles is analysed, which is directly correlated to the wear producing process in the tribo system. This is necessary because particle morphology (size, form factor) depends on the wear mode and characterizes the ongoing damage [3]. The particle size and form was measured and fotographed by SEM and the composition of the wear particles was analysed by EDX (micro probe). The results are shown in Figures 3, 4, 5 and 6. Subjected to the results of Debris Tests, during SEM-investigation attention was paid if a correlation of structures within the particles and the cause of wear could be found. Indications for adhesive wear (sliding wear, rolling wear), abrasive wear and spalling were detected. Particles were found with a size up to the order of 0.5x0.2x0.01mm. Particle features were round to oval, shaped like

tongues or scales, tiny and sharp edged and sometimes with striations. Particle morphology generally pointed to a damage due to fatigue, particle composition to alloys typically used for bearings. The elemental composition data obtained from EDX-analysis were classified according to particle size (Figures 7, 8). These figures illustrate that depending on the particle composition significant changes in particle size and form are found after 109.55 and 110.25 hours of operation due to a progress of damage. The detection of anomalous wear, the SEM/EDX-results and a detailed knowledge of the jet engine identify a failure of bearing No.6 in the rear bearing chamber.

Pattern recognition

The preceding discussion of a failure analysis of a jet engine shows, that a lot of expert knowledge is necessary, to extract essential information available from analytical data, so it can be used for condition monitoring and to specify the failing component. Expert knowledge about the construction and operational behavior of the jet engine has to be combined with expert knowledge about tribology and analytical techniques. It is obvious that this approach is only possible in special cases, but not on a routine basis. In our laboratory about 5000 jet engines and gear boxes of 40 different types are under early failure detection control. An increasing number of units is equipped with magnetic plugs and so did the number of wear particles to identify. Additionally depending on the system up to 30 different relevant alloys are used of which particles are found in the debris. Having to identify a lot of particles imposes the necessity to automate the wear particle characterization. In order to improve condition monitoring and to make expert knowledge available for routine work we are investigating the potentials of hierarchical cluster analysis (CA) and principal component analysis (PCA). The intention is to incorporate multivariate statistics as an integral part of early failure detection under the aspect of an automated failure pattern recognition which can be expanded to an expert system. Basically two main applications are offered: a.) an automated alloy identification; b.) the detection of characteristic patterns in the data, which can be related to a certain failure mode. Theory of CA and PCA has been explained in detail in [4, 5].
CA uses a mathematical algorithm to join together n-dimensional data sets into successively larger groups on a similarity scale.The result is a dendrogram (hierarchical tree) which shows the connection between the data sets based on their distance in n-space. In the application dicussed here, n-space is spanned by the element coordinates of the composition of reference alloys or wear particles. A data set can also include other wear relevant information like particle features (size, form factor, Debris units) and vibration analysis data (frequencies, amplitudes). For data analysis we generally use Euclidean or z-transformed distances as distance metric. In most cases the agglomerative methode used for hierarchical cluster analysis is uncritical. Best results were obtained with "Lance and Williams flexible" [6] or the "centroid"-method. Figure 9 shows the results of the cluster analysis of data matrices including data about the elemental composition and the size of wear particles and reference alloys. The numbers at the x-axis refer to the data set number of the particle to identify or reference alloys in the data matrix. The data belong to the failure case discussed before. The 4 dendrograms describe the wear

179

debris collected at the magnetic plugs after 109.55 h operation and an additional test run. To simplify the dendrogram, the data matrices only contain a reduced number of reference alloys and not the whole set necessary for routine wear particle identification. The different branches of the dendrograms separate the data depending on their similarity. Clustering dominantly is caused by the elemental composition of the particles to identify. While it is timeconsuming and complicated to compare numeric reference data with a large number of wear particle compositions directly, the graphical form to display the data makes it fast and easy to extract the essential information. Additionally, the comparison of numerical data is hindered because of the scatter of the analytical results. In contrary the comparison in n-space clusters according to the overall composition and gives a better discrimination between resembling alloys. The particles identified can be related as discussed before to the outer race and roller of bearing No. 6 and the Ni-coating of a drive shaft or sealing ring. The main branches of the dendrogram separating different alloys are split up into subgroups. This splitting is caused by a lower similarity of the data according to the size of the wear particles produced. Increasing operation time of the jet engine results in an increased number of larger particles, indicating the dramatic development of the failure of the ball bearing. The percentage of Ni-containing particles is reduced. While a dendrogram is best suited for the simple identification of the alloy of a wear particle by the position of the data point, its interpretation tends to be complicated if more information is included in the data set. Detailed information of a failure pattern can be obtained by principal component analysis (PCA) of the data set. PCA is a linear algebra technique which attemps to describe the quantity of the observed data by a smaller number of underlying factors. The abstract factors (eigenvectors or principal components) are used to visualize the n-dimensional measurement space by projecting the data sets down onto the first few eigenvectors. This projection gives a two or three dimensional view of the data, preserving as much of the original information as possible. This fraction of the original information is quantified in the "scores plot", which can be used to perceive obvious groupings (patterns) among the data sets. Figure 10 shows a scores plot of a data matrix after PCA. Because the matrix contained data about composition and the size of wear particles, the diagram corresponds to a certain configuration of particles collected. This configuration can be typical for a failure pattern. In the diagram three main clusters are visible, which are caused by the different wear particle species produced. The pattern is representative for the failure of ball bearing 6. So with a set of characteristic failure patterns even an unexperienced analyst is put in position to identify a special failure mode by comparing different graphs.

CONCLUSION

Condition monitoring of stressed machine parts can be obtained from analytical data and size distributions of debris particles generated in an oil-wetted circuit and released in oil. The combination of wear debris characterization and mathematical pattern recognition techniques aids to automate condition monitoring and can be expanded to an expert system. The diagnostic and prognostic capabilities of this technique will reduce the need for experts and will in spite of that

provide advanced condition monitoring prior to a mechanical system component failure.

SUMMARY

The example demonstrates the feasibility of using debris test and SEM- and EDX-analysis of wear particles combined with hierarchical cluster analysis and principal component analysis to nonintrusively monitor key mechanical components without expert knowledge. Such kind of m nitoring can reduce time consuming investigations and/or provide additional information concerning the correlation of changes in wear rate and changes in operating conditions. The cost of wear damage will continue to provide motivation for developing and improving methods for wear measurement and computer based techniques. The combination of computer expertise with the results gained from fundamental measurement methods can lead to successful data base structuring in conjunction with an understanding in physical behavior of tribosystems in jet engines. In contrary to numerical data, the graphical display of the results (particle size, formfactor, composition, operation time, cluster analysis, etc.) obtained by the methods mentioned before shows easy to understand wear profils and how degradation in the jet engine will result in a shift in this wear profile. These graphical data early indicate the degree of an engine damage and the wear mode, point to the location of a damage and help to jecide about necessary maintenance.

ACKNOWLEDGEMENT

REFERENCES

1. R.V.Kadyszewski, Advanced Quantitative Debris Monitoring, Debris Data Processor A Tool For Enhanced Weapon System 1986, Glenolden, PA, Cambridge University Press

2. M.H.Jones, J.Guttenberger, H.Brenneke, Condition Monitoring 1991 Erding, Pineridge Press, Swansea, UK

3. W.Uedelhoven, W.Mayr, M.Vitzthum, J.Guttenberger Condition Monitoring 1991, Erding, Pineridge Press, Swansea, UK

4. E.R.Malinowski, D.C.Howery, Factor Analysis in Chemistry, Wiley, New York (1980)

5. M.A.Scharaf, D.L.Illmann, B.R.Kowalski, Chemometrics, Wiley, New York (1986)

6. G.H.Lance a. W.T.Williams, A general theory of classificatory sorting strategies I, Hierarchical systems, Comp.J. 9 (1966), 373-380

Table 1: Materials of oil-lubricated components

Magnetic plug black (5), module 16 (3 and 4)	
Material	Description
8 CrMoVW 16 5 20 24 X 20 WCr 103 9 CrVW 12 18 8 100 Cr 6 X 12 CrNiMo 12	inner-bearing outer-bearing rings, rollers, balls bearing cages labyrinth-housing

Table 2: Materials of oil lubricated components

Magnetic plug orange (4), module 8, 10 and 11	
Material	Description
X 12 CrNiMo 12 X 20 WCr 103 9 CrVW 12 18 8 100 Cr 6 IN 718 (Cr19Mo3Ni52Fe19) X 20 WCr 103 IN 718 9 CrVW 12 18 8 100 Cr 6 X 12 CrNiMo 12 9 CrVW 12 18 8 100 Cr 6	seal-labyrinth tube inner ring-bearing (No. 5) bearing-roller (No. 5) bearing-cage (No. 5) flange shaft outer ring-bearing (No. 5) inner-seal outer- and inner ring, balls (No. 7) bearing cage (No. 7) ring-seal outer-ring and roller (No. 6) retainer-roller

Table 3. Operating time of the jet engine and debris units (DTU)

date	operating time (h)	specimen No.	magnetic plugs					
			designation					
			white DTU	green DTU	yellow DTU	orange DTU	black DTU	blue DTU
1.2.	24	2	7	7	21	15	11	13
24.2.	37	3	20	12	6	13	6	14
8.3.	50	4	2	3	5	7	3	10
22.3.	62	5	4	3	7	6	4	4
3.4.	75	6	2	3	4	6	3	9
17.4.	88	7	5	6	7	9	9	7
9.5.	101	8	4	3	6	9	4	9
6.8.	109	9	240	415	750	1911	695	250

R B 199

magnetic plug white (1)   magnetic plug blue (6)   magnetic plug
module 1,2   module 3,4   orange (4)
                                                                     module 8,10,11

                           magnetic plug green (2)
                           magnetic plug yellow (3)   module 3,4,16
                           magnetic plug black (5)
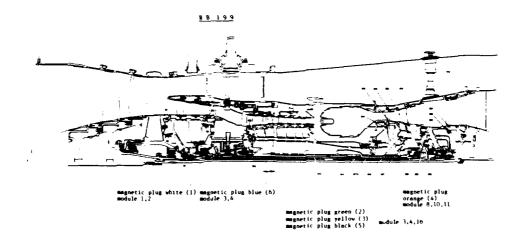
Figure 1
Cross-section of a modern jet engine; the figure illustrates areas where
critical components are located
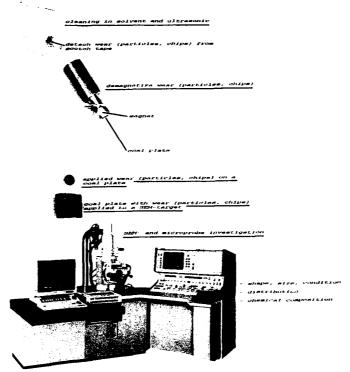
Figure 2
The illustration shows how wear was prepared for SEM/EDX-investigations
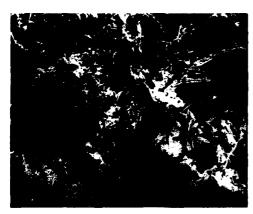
Figure 3
SEM photograph showing numbers of
typical particles collected and
identified (magn.plug black 110h)

Figure 4
SEM photograph of wear particles
with different form and size
(magn.plug orange 110h)

184
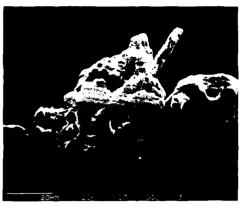
Figure 5
SEM photograph showing Ni-particles
(magn.plug orange
110.25 hours operating time)
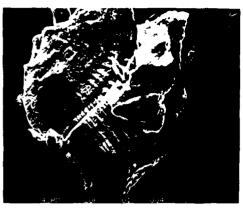
Figure 6
Significant particle illustrates
the form due to spalling wear
(magn.plug orange 110.25h)
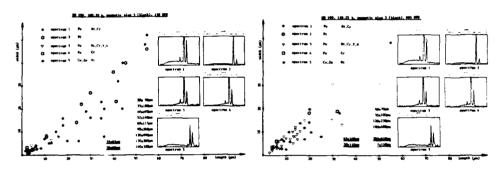
Figure 7
The data show the distribution of wear particle size (length x-axis,
width y-axis) and composition (magnetic plug black, 109.55 h operating
time)
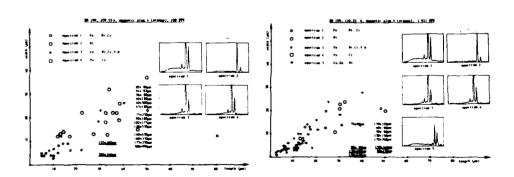
Figure 8
These data show the distribution of wear particle size and composition
in comparison to Figure 7 (magn.plug orange 110.25h)

185
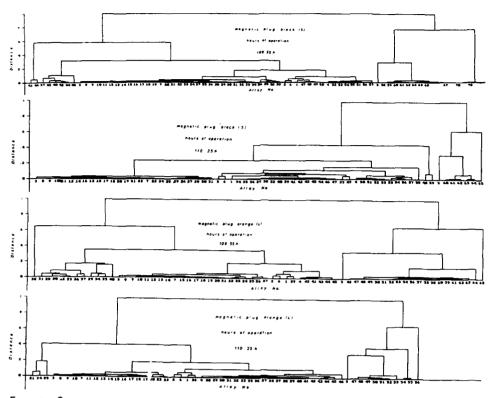
Figure 9
Dendrograms showing the difference between magnetic plug black (109.55h
and 110.25h) and magnetic plug orange (109.55h, 110.25h of operation)
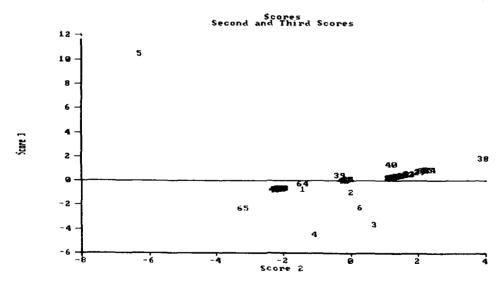
Figure 10
Score plot. Typical pattern for a fatigue failure of bearing No.6

186

SYNCHRONOUS SIGNAL PROCESSING TECHNIQUES FOR BEARINGS AND OTHER MACHINERY COMPONENTS

Walter Hernandez, Ph.D
Monitoring Technology Corporation
Falls Church, VA 22043

Abstract: Today's vibration monitoring systems for machine fault detection tend to use modern techniques like pattern recognition and expert systems for decision analysis, but use outdated signal processing techniques for the front-end of the system. The signal processing usually consists of forming FFT power spectra and shaft signal averages. Developed prior to 1975, these techniques alone are inadequate for analyzing complex machinery like helicopter gearboxes. Shaft signal averaging, a synchronous technique, has proven useful in detecting gear faults. But most investigators feel this synchronous technique, while useful for gears, is limited in its applicability. However, MTC, over the last eight years, has developed a set of synchronous processing techniques which are generally applicable to complex machines and a wider range of components such as gears, bearings, pulleys, and blades. These techniques detect and separate the signatures of each of the machine's components based on the coherent properties of the components. Several of these techniques with application examples are presented in this paper.

Key Words: Vibration; detection; faults; gears; bearings; signal processing; rotating machinery

Introduction: Computer based vibration monitoring systems are in a rapid state of development. A main reason for this activity is the promise of predictive maintenance, i.e., the early detection of developing faults in machinery. This results in maintenance cost savings, the increased up-time and availability of machinery, and, finally the increased safety of properly monitored machinery.

In this paper, we shall discuss the basic vibration monitoring system as it might be applied to a complex rotating machine and, in particular, complex transmissions. We show that new signal processing techniques, many which are probably unfamiliar to the reader, are now available and can greatly increase the reliability and effectiveness of these monitoring systems.

Basic Vibration Monitoring System: Figure 1 shows the basic parts of a conventional on-line vibration monitoring system for fault detection of a complex rotating machine such as a helicopter main gearbox. We show three main elements: 1) the gearbox; 2) the signal processing element; and, 3) the decision analysis element which yields the current fault status of the gearbox.

The gearbox illustrated shows two engine input shafts, one output shaft, and indicates that there could be up to 50 gears and 100 bearings present in this reduction gear drive. Also shown are three vibration sensors (there could be many more) and one encoder sensor which often is simply a tachometer type device which magnetically senses each rotation of the shaft.

The signal processing element of this conventional system collects the signals from each of the sensors, performs analog signal conditioning, digitizes the vibration signals and digitally generates FFT type power spectra for each sensor. Also generated, with the aid of the tachometer pulse, are time domain signal averages for each gear (or shaft) of the drive train.

Finally, the decision analysis element gathers these signatures (the power spectra and signal averages) and analyzes them to determine if any faults are developing. Techniques such as feature (FOM) extraction, trending and alarming, plus more advanced A.I. methods like expert systems and neural networks are currently being developed and applied.

Problem: Inadequate Signal Processing: A complicated gearbox produces a complex, noisy vibration signal which cannot be reliably analyzed with currently used signal processing techniques, techniques which were developed largely, in fact, prior to 1975. Figure 2 is an example of this for the case of a vibration power spectrum from reduction GEARBOX A (see Figure 4). The range of this spectrum is the narrow band of 800-826 Hz, yet the complexity is staggering. The myriad of spectral lines shown in this high resolution spectrum are mainly due to shaft interactions which modulate the 813 Hz gearbox mesh. Some noise components are also indicated. Bearing lines are also present, but, as we shall see, difficult to detect and classify.

Solution: Synchronous Signal Processing (SSP): SSP consists of a body of hardware and software techniques which enables the processor to separate the signatures of the various machine components. These techniques, based on the synchronous or coherency properties of the machine and employed on today's fast and inexpensive computers, yield startling improvements in component signature analysis. Eight (8) SSP techniques listed below are discussed herein. Three of these techniques, shaft encoders, array processors, and discrete Fourier Transforms, have been in use for some time but are reviewed for completeness. The remaining five techniques were developed by MTC (1985-1993).

- Shaft Encoders (hardware)
- Multiple Top Dead Centers (hardware)

- Array Processors (hardware)
- Discrete Fourier Transforms (software)
- Hunting Tooth Segmentation (software)
- Hunting Tooth Averages (software)
- High Frequency Averages (software and hardware)
- 2-Form Spectra (software)

SSP Technique/Shaft Encoders (Hardware): A shaft encoder is a device which produces a set of uniformly spaced pulses per shaft rotation. There are two great advantages afforded by these devices. One, they can be used as a clock to control the A/D process of the vibration signals. This yields a fixed number of samples per shaft rotation enabling order analyses and the removal of RPM variation effects. Figure 3 shows a hypothetical power spectrum of a rotating machine with RPM variation using conventional interior clock A/D control and exterior clock (encoders) A/D control. The well defined spectral orders and magnitudes of the latter are apparent.

A second use of the encoder is the direct analysis of the encoder pulses themselves. This can yield valuable information in the form of FM and AM detection of machine vibration, e.g., turbine blade analysis. This will not be discussed here.

SSP Technique/Shaft Top Dead Centers (TDC Hardware): A shaft top dead center is a device which produces a single pulse for each complete rotation of that shaft. If several of these devices are properly placed on complex gearboxes -- often two is sufficient -- one is able to uniquely determine the angular orientation of every gear and shaft in the gearbox. An example of the advantages this yields is illustrated in Figure 4 for GEARBOX A. Here, a TDC on the input and output shafts allows one to generate gear signal averages which are perfectly aligned by tooth number. This is accomplished by beginning the contiguous averaging processing when pulses from the two TDCs are coincident. Thus, one can determine if apparent tooth faults detected at different times are really the same tooth or not the same tooth. We call this tooth tracking.

SSP Technique/Array Processors (Hardware): An Array Processor (AP) is a digital processor with specialized hardware architecture designed to achieve rapid throughputs for large vector operations. These devices, coming into common use in audio, video, and communications fields, can accomplish high speed calculations at low costs. Figure 5 compares the relative speeds of the calculation of spectra using the DFT algorithm vs. the FFT algorithm in 1965 and 1993 (using an AP). We shall see in the next section that there are great advantages to using the DFT compared to the FFT, but in 1965 computers were so slow that the Cooley-Tukey FFT Algorithm was the only practical spectrum

computation method. But in 1993, this is no longer the case. Using an AP, the DFT, with all its advantages, can regain its preeminent status.

SSP Technique/Discrete Fourier Transform (Hardware): The discrete Fourier transform is the Fourier transform of a digital signal of N discrete points where N is an arbitrary integer. In contrast, the commonly used Fast Fourier Transform (FFT) is restricted to data lengths equal to a power of 2, e.g., 512, 1024, 2048. The most important advantage of the DFT is that data lengths can be selected for specific machines so that important spectral components of that machine can be made to fall exactly on the discreet spectral values calculated by the DFT. Figure 6 illustrates this by comparing typical FFT and DFT spectra for a single spectral component. By using the DFT, leakage spreading of spectral energy to nearby spectral values is eliminated. This enables the exact identification of spectral components and the measurement of their exact amplitudes.

SSP Technique/Hunting Tooth Segmentation (Software): HT segmentation is the division of data into contiguous time segments equal to the cycle time of the gearbox. Figure 7 illustrates this for the GEARBOX A discussed earlier. For the gear tooth numbers indicated, the cycle time is equal to 17x71 turns of the input shaft. (Because the meshing gears of 25 teeth and 85 teeth have a common factor 5, the digit 85/5 = 17 is used in the HT calculation.) For an encoder generating 145 pulses per input shaft rotation, we have the HT as 175,015 data samples! The advantage of performing HT spectra with the DFT as compared to common FFT spectra is shown. Spectral components which appear as several peaks in the FFT reveal themselves as a complex, highly defined spectrum of interaction between the meshing frequency at 813.314 HZ and the various rotating shafts. (This we see is the source of Figure 2.) This technique allows the exact identification of the gearbox spectral components and their exact amplitude measurements. *Note*, the HT can be very long and the averaging of these types of records can require very large amounts of data. However, for on-line systems, the data is available.

SSP Technique/HT Signal Averaging (Software): HT signal averaging is the synchronous averaging of vibration data using the HT period instead of the commonly used shaft periods. There are many advantages to this technique including 1) more sensitive tooth defect detection, 2) faulty gear identification for gears which have identical rotation rates and 3) elimination of "apparent" defects. Figure 8 shows an example of a HT average for a pinion and wheel truck axle. Four large spikes in the HT average all occur when tooth #8 of the pinion is engaged, leaving little doubt that an anomaly exists on tooth 8. The shaft averages are also shown. Tooth #8 of the pinion still

190

shows a maximum value, but its magnitude relative to the other teeth has been reduced in comparison to the HT result. The wheel gear shows 4 spikes that we call "apparent" defects since we know the defect is actually on the pinion. With regard to these shortcomings, the superiority of the HT average is evident.

SSP Technique/High Frequency >20 KHZ Averaging (Hardware and Software): In this technique, the signal average is performed using the energy or envelope of a high frequency band. The main advantage here is that high frequency vibration data contains valuable fault information and often the S/N is superior to the low frequency bands. Figure 9 shows an example of a tooth crack developed on a fatigue test stand. The left hand plot shows the standard shaft averaging results at times T1, T2, and T3 as the crack develops, whereas the right hand plot shows the result of averaging the envelope of the 30-50 KHZ band. The crack is detected very distinctly in the HF data by the appearance of 2 large teeth. The crack has failed to show at all in the standard shaft average. Note that standard shaft averaging usually eliminates frequency components greater than 10 KHZ.

SSP Technique/2-Form Spectrum (Software): A 2-Form Spectrum is the multiplication of non-synchronous spectral components to produce synchronous components. The advantage of 2-Form Spectrum is that it can perform synchronous detection of bearings, blade resonances, and other non-synchronous type machine components. In addition, it can separate these components from other synchronous and non-synchronous components in the spectrum. Figure 10 shows an example of this technique applied to the GEARBOX A discussed earlier. The top plot shows the HT/DFT/Power Spectrum. There is a lot of detail, but the bearing lines cannot be clearly identified. The bottom plot shows the results of calculating a HT/DFT/2-Form, i.e., replacing the power spectrum with a 2-Form spectrum and also with automatic shaft line deletion. The bearing cage modulation component about the gear mesh is now readily evident. All shaft and noise components are eliminated.

Current SSP Users: Several of these SSP techniques are available in MTC's G-3000 system and are being used in a variety of applications. These include:

- End-of-line testing of superchargers for Eaton Corporation

- Engineering testing of nose gearboxes for Pratt & Whitney, Inc.

191

- Predictive maintenance of cement mill gear drives for MAAG Gear Co., Ltd. of Switzerland

- Fault detection in gear pair fatigue testing for Sundstrand Corp., NASA Lewis, and Rockwell Int'l.

Summary: Several points are to be emphasized:

- Standard signal processing techniques are inadequate for detecting and separating component signatures in complex machinery such as gearboxes used in helicopters, tanks, and destroyers.

- New synchronous signal averaging techniques designed for rotating machinery and developed at MTC over the last eight years are very effective in detecting and separating component signatures of complex machinery.

- Advanced A.I. type technology, such as expert systems and neural networks, cannot compensate for poor signal processing. Good signal processing, on the other hand, will make these A.I. techniques more effective.

VIBRATION SENSORS
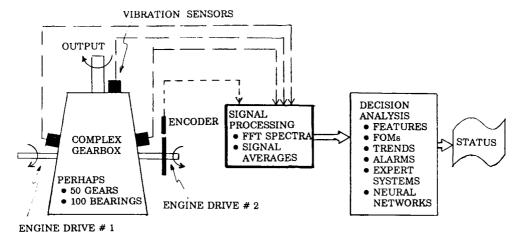
FIG. 1 CONVENTIONAL VIBRATION MONITORING SYSTEM
FOR ON-LINE GEARBOX FAULT DETECTION
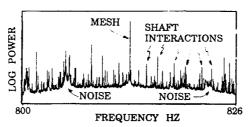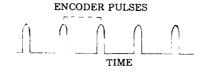
FIG. 2 NARROW-BAND HIGH RESOLUTION
POWER SPECTRUM OF GEARBOX A

ENCODER PULSES

TIME
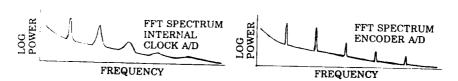
FIG.3   THE UPPER PLOT SHOWS PULSES FROM A SHAFT
ENCODER. THE LOWER LEFT PLOT SHOWS THE
COMMON   FFT SPECTRUM FOR A MACHINE WITH
VARYING RPM. THE LOWER RIGHT PLOT SHOWS THE
SPECTRUM OF THE SAME MACHINE WITH AN
ENCODER DRIVEN A/D.

193

GEARBOX A

SHAFT AVERAGE W/O TDCs

SHAFT AVERAGE WITH TDCs

1 REV 25T GEAR

1 REV 25T GEAR
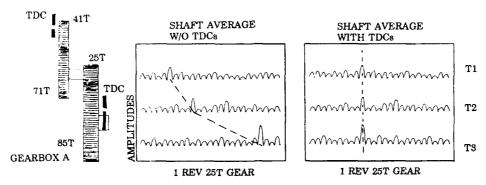
FIG.4   COMPARISON OF SHAFT AVERAGES WITH AND WITHOUT MULTIPLE TDCs ILLUSTRATES TOOTH TRACKING AT TIMES T1, T2, AND T3

		1965	1993
DFT	1024 PTS	~1 HR	~.2 SEC
FFT	1024 PTS	~ 25 SEC	~ .002 SEC
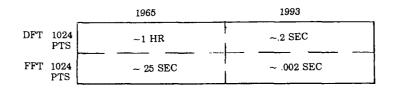
FIG. 5   COMPARISON OF DFT AND FFT COMPUTATION TIMES IN 1965 AND 1993
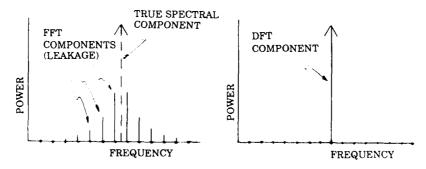
FIG. 6   COMPARISON OF FFT SPECTRUM AND PROPERLY CHOSEN DFT SPECTRUM FOR A SINGLE SPECTRAL COMPONENT

194
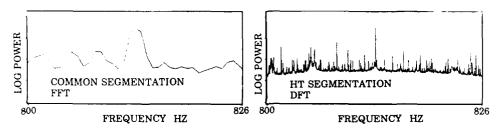
FIG 7   COMPARISON OF STANDARD FFT SEGMENTATION
         SPECTRUM (4096 PTS) AND HT SEGMENTATION
         DFT SPECTRUM (175,015 PTS)
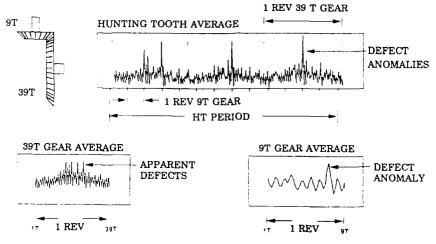
FIG. 8   COMPARISON OF HUNTING TOOTH AVERAGE
          TO EACH OF THE SHAFT AVERAGES FOR
          SINGLE DEFECT ON PINION GEAR
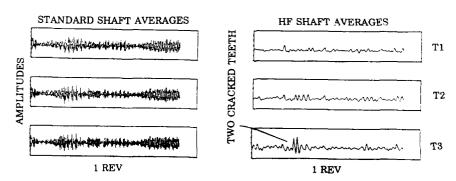
FIG. 9   COMPARISON OF STANDARD SHAFT AVERAGES
          AND HIGH FREQUENCY SHAFT AVERAGES FOR
          TOOTH FAULT DETECTION AT TIMES T1, T2, T3

195
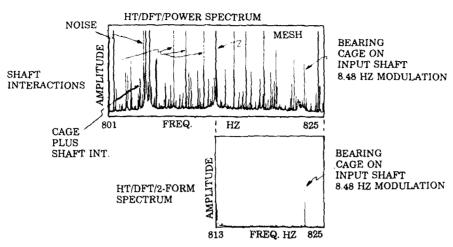
FIG. 10   THE HT/DFT/POWER SPECTRUM IS COMPARED
TO THE HT/DFT/2-FORM SPECTRUM (WITH
DELETION OF SHAFT INTERACTIONS).

196

NONDESTRUCTIVE EVALUATION
AND
INFORMATION PROCESSING II

Cochairmen: Paul L. Howard
              Paul Howard Enterprises

              William Prosser
              NASA Langley Research Center

AN ANALYSIS OF GEAR FAULT DETECTION METHODS AS APPLIED TO PITTING FATIGUE FAILURE DATA

James J. Zakrajsek and Dennis P. Townsend
National Aeronautics and Space Administration
Lewis Research Center
21000 Brookpark Road
Cleveland, Ohio 44135

Harry J. Decker
Propulsion Directorate
U.S. Army Aviation Systems Command
Lewis Research Center
21000 Brookpark Road
Cleveland, Ohio 44135

Abstract: The application of gear fault prediction techniques to experimental data is examined. A single mesh spur gear fatigue rig was used to produce naturally occurring faults on a number of test gear sets. Gear tooth surface pitting was the primary failure mode for a majority of the test runs. The damage ranged from moderate pitting on two teeth in one test to spalling on several teeth in another test. Previously published failure prediction techniques were applied to the data as it was acquired to provide a means of monitoring the test and stopping it when a failure was suspected. A newly developed technique along with variations of published methods were also applied to the experimental data. The published methods experienced some success in detecting initial pitting before it progressed to affect the overall root-mean-square (RMS) vibration level. The new technique robustly detected the damage on all of the tests, and in most cases continued to react to the damage as it spread and increased in severity. Since no single method was able to consistently predict the damage first on all the runs, it was concluded that the best approach to reliably detect pitting damage is to use a combination of detection methods.

Key Words: Gear; Fatigue; Diagnostics; Failure Prediction

Introduction: Drive train diagnostics is becoming one of the most significant areas of research in rotorcraft propulsion. The need for a reliable health and usage monitoring system for the propulsion system can be seen by reviewing some rotorcraft accident statistics. An investigation of serious rotorcraft accidents that were a result of fatigue failures showed that 32 percent were due to engine and transmission components [1]. Also, 60 percent of the serious rotorcraft accidents were found to occur during cruise flight. Civil helicopters need a thirtyfold increase in their safety record to equal that of conventional fixed-wing turbojet aircraft. Practically, this can only be accomplished with the aid of a highly reliable, on-line health and usage monitoring unit. Diagnostic research is required to develop and prove various fault detection concepts and methodologies.

A number of methods have been developed to provide early detection of gear tooth surface damage. McFadden proposed a method to detect gear tooth cracks and spalls by demodulating the time signal [3]. Stewart devised several time domain discriminant methods

of which FM0, a coarse fault detection parameter, and FM4, an isolated fault detection parameter, are the most widely referenced [4]. Martin proposed using the sixth and eighth statistical moments of the time signal to detect surface damage [2]. A new method, NA4, was recently developed at NASA Lewis Research Center to detect and continue to react to gear tooth surface damage as it spreads and grows in severity.

Verification of these detection methods with experimental data along with a comparison of their relative performance is a crucial step in the overall process of developing a highly reliable health monitoring system.

In view of the aforementioned, it becomes the object of the research reported herein to determine the relative performance of the detection methods as they are applied to experimental data. Each method is applied to vibration data obtained from a gear fatigue test rig at NASA Lewis, where test gears are run until a fatigue failure occurs. The failure modes of the five tests used in this study ranged from moderate pitting on two teeth in one test to spalling on several teeth in another test. Results of each method are compared for each test, and overall conclusions are made regarding the performance of the methods.

Theory of Fault Detection Methods: All of the methods in this investigation utilized vibration data that was preprocessed as it was collected. To eliminate the noise and vibration that is incoherent with the rotational speed of the test gears, the raw vibration data was time synchronous averaged immediately after being digitized. During time synchronous averaging, the data was also interpolated to obtain 1024 points per revolution of the test gears. Each of the methods presented below were then applied to the time averaged and interpolated vibration data.

FM0 is formulated to be a robust indicator of major faults in a gear mesh by detecting major changes in the meshing pattern [4]. FM0 is found by dividing the peak-to-peak level of the signal average by the sum of the amplitudes of the mesh frequency and its harmonics. In major tooth faults, such as breakage, the peak-to-peak level tends to increase, resulting in FM0 increasing. For heavy distributed wear or damage, the peak-to-peak remains somewhat constant but the meshing frequency levels tend to decrease, resulting in FM0 increasing.

FM4 was developed to detect changes in the vibration pattern resulting from damage on a limited number of teeth [4]. A difference signal is first constructed by removing the regular meshing components (shaft frequency and harmonics, primary meshing frequency and harmonics along with their first order sidebands) from the original signal. The fourth normalized statistical moment (normalized kurtosis) is then applied to this difference signal. For a gear in good condition the difference signal would be primarily noise with a Gaussian amplitude distribution, resulting in a normalized kurtosis value of 3 (nondimensional). When one or two teeth develop a defect (such as a crack, pit, or spall) a peak or series of peaks appear in the difference signal, resulting in the normalized kurtosis value to increase beyond the nominal value of 3.

A demodulation technique was developed to detect local gear defects such as fatigue cracks, pits and spalls [3]. The basic theory behind this technique is that a gear tooth defect will produce sidebands that modulate the dominant meshing frequency. In this method, the signal is band-passed filtered about a dominant meshing frequency, including as many sidebands as possible. The Hilbert transform is then used to convert the real band-passed signal into a complex time signal, or analytic signal. The normalized kurtosis is then applied to the amplitude modulation function (magnitude of the analytic signal) in an attempt to

detect gear tooth damage through the modulating sidebands. Again, a value of 3 would indicate a nominal condition, and a value over 3 indicates possible tooth damage.

M6A and M8A are variations of the sixth (M6) and eighth (M8) normalized statistical moments proposed by Martin to detect surface damage using vibration signals [2]. M6 and M8 are applied to the same difference signal as defined in the definition of FM4. The basic theory behind M6A and M8A is the same as that for FM4, except M6A and M8A should be more sensitive to peaks in the difference signal. Also, the value for nominal conditions (Gaussian distribution) is 15 for M6A, and 105 for M8A.

NA4 is a new method that was developed by the authors to not only detect the onset of damage, as FM4 does, but also to continue to react to the damage as it spreads and increases in magnitude. Similar to FM4, a residual signal is constructed by removing regular meshing components from the original signal, however, for NA4, the first order sidebands stay in the residual signal. The fourth statistical moment of the residual signal is then divided by the current run time averaged variance of the residual signal, raised to the second power, resulting in the quasi-normalized kurtosis given below:

$$NA4(M) = \frac{N \sum\limits_{i=1}^{N} (r_i - \bar{r})^4}{\left\{ \dfrac{1}{M} \sum\limits_{j=1}^{M} \left[\sum\limits_{i=1}^{N} (r_{ij} - \bar{r}_j)^2 \right] \right\}^2}$$

where

r	residual signal
$\bar{r}$	mean value of residual signal
N	total number of data points in time record
i	data point number in time record
M	current time record number in run ensemble
j	time record number in run ensemble

In NA4, the kurtosis is normalized, however it is normalized using the variance of the residual signal averaged over the run up to point in the run that NA4 is being calculated for. With this method, the changes in the residual signal are constantly being compared to the running average of the variance of the system, or a weighted baseline for the specific system in "good" condition. This should allow NA4 to grow with the severity of the fault until the average of the variance itself changes. As with FM4, NA4 is dimensionless, with a value of 3 under nominal conditions.

Apparatus and Gear Damage Review: A spur gear fatigue rig at NASA Lewis was used to obtain experimental data. The primary purpose of this rig is to study the effects of gear materials, gear surface treatments, and lubrication types on the surface fatigue strength of aircraft quality gears. The rig was recently modified to allow it to be used for diagnostic studies as well as fatigue research [5]. Vibration data from an accelerometer mounted on a bearing end plate was captured using an on-line program running on a personal computer with an analog to digital conversion board and anti-aliasing filter. The test gears are standard spur gears having 28 teeth and a pitch diameter of 88.9 mm (3.50 in.). The gears were loaded to 74.6 Nm (660 in.-lb) at an operating speed of 10,000 rpm.

Some examples of the different magnitudes of tooth damage found in the five tests (runs 1 to 5) of this study are illustrated in Figure 1. Figure 1(a) shows the isolated heavy pitting damage that was found on the test gears in run 1 at 131 hours into the test. Figure 1(b) shows an example of the spalling damage found at the end of the test of run 1. Figure 1(c) illustrates an example of the moderate pitting found in the tests. Similarly, Figure 1(d) gives an example of the heavy pitting damage found in the tests. Details of the damage found in each test are given below, with Figure 1 serving as a pictorial reference of damage magnitude.

At 131 hours into run 1, damage was found on two teeth on the driver gear (one heavy and one moderate pitting). Both mating teeth on the driven gear were also found to be damaged (both heavy pitting). Figure 1(a) illustrates the heavy pitting damage on the driver and driven gears at 131 hours. At the end of run 1, spalling (Figure 1(b)) and heavy pitting damage was found on roughly one third of the teeth on both the driver and driven gears.

At the end of run 2, damage was found on three consecutive teeth on the driver gear (one heavy and two moderate pitting). Two of the three mating teeth on the driven gear were also found to be damaged (both moderate pitting).

At the end of run 3, damage was found on four consecutive teeth on the driver gear (one spalling, two heavy, and one moderate pitting). One of the four mating teeth on the driven gear was also found to be damaged (moderate pitting).

At the end of run 4, damage was found on two consecutive teeth on the driver gear (both heavy pitting). The two mating teeth on the driven gear were also found to be damaged (one heavy, and one moderate pitting).

At 294 hours into run 5, micropitting and wear was found on nearly all the teeth of the driver gear. At the end of run 5, moderate pitting was found on eight teeth distributed on the driver gear. Three consecutive teeth on the driven gear were found to have moderate pitting damage.

Discussion of Results: The results of applying the fault detection methods to the experimentally obtained vibration data are illustrated in Figures 2 to 6.

Figure 2 presents the results of all the parameters for run 1. The vertical centerline in each plot represents the point in time (t = 131 hours) in which the rig was stopped and the damage was recorded, as shown in Figure 1(a). As seen in Figure 2, the parameters FM4, NA4, Kurtosis of AMF, M6A, and M8A all detect tooth damage at t = 110 hours, or 25 hours before FM0 reacts, and 27 hours before the overall root-mean-square (RMS) vibration level increases. FM4 peaked at a value of 5.4, then dropped off to the nominal value of 3 at t = 131 hours. It is possible that only one of the two teeth found damaged at t = 131 hours actually started at the time FM4 reacted, and when the damage spread to the other tooth, FM4 lost its sensitivity by decreasing back to its nominal value. The results of the demodulation method for run 1 (Figure 2(e)), are the best results obtained from that method. In other runs it showed results very similar to FM4 results (runs 2 and 3), or gave no indication at all (runs 4 and 5). As seen in Figure 2, M6A and M8A results follow the same trends indicated by FM4. M6A and M8A, however reacted more strongly to the damage; as indicated by the 300 percent and 863 percent increases over nominal values for M6A and M8A, respectively, as compared to an 80 percent increase for FM4. These results for M6A and M8A are very typical of the results obtained for M6A and M8A on the other four runs. FM0 gave a solid indication of over three times its nominal value, and 2 hours in

advance of the RMS level increase. NA4 gave the best performance for run 1. Figure 2(f) shows the first 135 hours of Figure 2(d), with an expanded vertical scale, for clarity. As seen in these two figures, NA4 reacts very robustly to the start of damage, sharply increasing to a value of 25, and remains somewhat steady at a value of 15 even as the other parameters (FM4, M6A, etc.) drop back down to nominal values. NA4 then increases sharply to a peak value of 230, following a trend similar to the RMS level increase. This could be the point at which the extremely heavy damage started (as seen in Figure 1(b)), continuing to the end of the run.

As seen in Figure 3, the parameters FM0, FM4, and NA4 all react sharply to the tooth damage at 94 hours into run 2. FM0 reacted robustly to the damage, increasing to over double its nominal value, whereas the overall RMS vibration level gradually increases with run time. FM4 also reacted by increasing from a value a little under the nominal 3 to a relatively steady value of 4.5 through to the end of the run. Because the heavy pitting damage was still isolated to only one of the three damaged teeth on the driver, FM4 was able to continue to react to the damage. NA4 gave the most robust reaction to the damage, increasing sharply from the nominal value of 3 to a value of 9 at t = 94 hours. NA4 then continues to increase from 9 to a peak of 29, growing gradually with the damage until 2 hours before the end where NA4 then drops off, due to a sharp increase in the denominator of NA4.

In run 3, only FM0 showed any significant reaction to the start of damage at 43 hours into the run, as seen in Figure 4. FM0 increased to over double its nominal value at this time. The damage may have been too subtle for the overall RMS level to increase, and may have started somewhat simultaneously over the four driver teeth for FM4 to indicate only a low grade response at t = 43 hours. When FM4 and NA4 do react at t = 74 hours, possibly due to the spalling formation on one of the four driver teeth, NA4 again reacts more robustly, increasing to 8, as compared to 5 for FM4. Both parameters increase, but FM4 peaks at 7.5, whereas NA4 peaks at 43.

As illustrated in Figure 5, the damage in run 4 was detected by FM0 and FM4 at the same time that the overall RMS vibration level increased. FM0 again shows a significant reaction to the damage, increasing in value to nearly three times its nominal value, as compared to the RMS level, which increases only 40 percent over its nominal value. FM4 peaks at 4.6, then proceeds to fall back to the nominal value. One of the two heavily damaged teeth on the driver gear may have started first, followed by heavy damage on the second tooth and the resulting decrease in the response and thus sensitivity of FM4. NA4 gives a strong indication of damage nearly 5 hours before the other parameters, and peaks at the value of 18.5, as compared to 4.6 for FM4. NA4 then decreases after the peak to 6.5, as its denominator increases, at the end of the run.

The vertical centerline in all the plots in Figure 6 indicate the point in time (t = 294 hours) that run 5 was stopped and micropitting was found on nearly all the teeth on the driver gear. As seen in Figure 6, FM0 and NA4 clearly detect the micropitting damage. After this point, FM0 and NA4 increase sharply, with FM0 peaking at over twice its nominal value, and NA4 increasing to a value of 15, then slowly dropping off. The sharp increase seen in FM0, NA4, and even the overall RMS vibration level most probably corresponds to the initiation of the moderate pitting found on a number of teeth on both driver and driven gears at the end of the run. As evident in Figure 6(b), FM4 gave no indication of either the initial micropitting damage nor the moderate pitting damage found at the end of the run. Due to the nature of the damage, both the micropitting and moderate pitting damage may

have occurred simultaneously on more than one or two isolated teeth, FM4 was incapable of reacting to it.

Based on the results just presented, it is clearly evident that of all the methods investigated in this study, the previously published method FM0 and the newly developed method NA4 are the most robust and reliable indicators of gear tooth pitting fatigue damage. FM0 gave a clear indication of the pitting fatigue damage on all five runs. On an average, FM0 increased to nearly three times its nominal value several hours before the RMS level showed any real change, on a majority of the runs. NA4 also gave a clear indication of the pitting fatigue damage on all five runs. NA4 reacted not only to isolated pitting damage on one or two teeth, but also to pitting damage that occurred over a number of teeth around the gear. NA4 gave robust initial reactions to the damage, increasing from the nominal value of 3 to an average value of 15, and in some cases continued to react as the damage spread and/or increased in severity.

The other methods were able to predict the pitting damage in most of the runs, however, they did not perform as reliably or robustly as FM0 and NA4. FM4 is a relatively good indicator of damage on one or two isolated teeth, however, results showed that as the damage spread to other teeth FM4 lost its sensitivity and dropped back down to the nominal value of 3. In one case FM4 never reacted, as the damage may have initiated on a number of teeth at approximately the same time. Although M6A and M8A showed stronger reactions to the damage, as compared to FM4, they exhibited the same trends as FM4, and thus the same weaknesses. The demodulation method gave results no better than FM4 in three of the runs, and failed to react to the damage in the remaining two runs.

In order to accurately and reliably detect gear tooth pitting fatigue damage, several methods need to be used in combination. Even with the limited data used in this study, not one method was able to give a first indication of the damage consistently on all five runs. Several methods, FM0 and NA4 as a minimum, need to operate in parallel in order to provide a reliable way of detecting the pitting damage as far in advance of severe damage as possible.

Conclusions: Based on this investigation, the following conclusions can be made

1) The newly developed parameter, NA4, reacted very robustly to the damage on all the runs. It reacted to isolated pitting damage as well as pitting damage on a number of teeth distributed around the gear. In several cases, NA4 continued to react as the damage spread and/or increased in severity, thus indicating damage level.

2) FM0 is a strong indicator of gear tooth pitting damage occurring over a number of teeth on a gear. For a majority of the runs, FM0 reacted to the damage before the RMS vibration level reacted. On those runs where FM0 reacted at the same time as the RMS level, FM0 gave a much clearer indication.

3) FM4 reacts well to damage on one or two isolated teeth, but loses its sensitivity significantly as the damage spreads to other teeth. FM4 failed to detect damage on one run as the pitting damage may have initiated on several teeth at the same period in time.

4) M6A and M8A exhibited stronger reactions to the damage, as compared to FM4. They, however, showed the same trends, and thus the same weaknesses, as FM4.

5) No single method was able to consistently predict the pitting damage before the others on all the runs. A number of the methods, FM0 and NA4 as a minimum, need to be used in combination in order to reliably detect gear tooth pitting damage.

204

REFERENCES

1. Astridge, D.G.: Helicopter Transmissions—Design for Safety and Reliability. Inst. Mech. Eng. Proc., Pt. G—J. Aerosp. Eng., vol. 203, no. G2, 1989, pp. 123-138.

2. Martin, H.R.: Statistical Moment Analysis As a Means of Surface Damage Detection. Proceedings of the 7th International Modal Analysis Conference, Society for Experimental Mechanics, Schenectady, NY, 1989, pp. 1016-1021.

3. McFadden, P.D.: Detecting Fatigue Cracks in Gears by Amplitude and Phase Demodulation of the Meshing Vibration. J. Vib. Acoust. Stress Reliab. Design, vol. 108, no. 2, Apr. 1986, pp. 165-170.

4. Stewart, R.M.: Some Useful Data Analysis Techniques for Gearbox Diagnostics. Report MHM/R/10/77, Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton, July 1977.

5. Zakrajsek, J.J., et al.: Analysis and Modification of a Single-Mesh Gear Fatigue Rig for Use in Diagnostic Studies. NASA TM-105416, 1992.

(a) Heavy pitting on two teeth in Run 1 at t = 131 hr into run.

(b) Example of spalling on tooth in Run 1 at end of test (t = 198 hr).

(c) Example of moderate pitting (Run 5, end of test).

(d) Example of heavy pitting (Run 3, end of test).

Figure 1.—Examples of actual damage on gear teeth.

(a) RMS.

(b) FM4.

(c) FMO.

(d) NA4.

(e) Kurtosis of amplitude modulation function.

(f) NA4 up to t = 135 hr..

(g) M6A.

(h) M8A.

Figure 2.—Run 1 results.

(a) RMS level.

(b) FM4.

(c) FMO.

(d) NA4.

Figure 3.—Run 2 results.

(a) RMS level.

(b) FM4.

(c) FM0.

(d) NA4.

Figure 4.—Run 3 results.

207

(a) RMS level.

(b) FM4.

(c) FMO.

(d) NA4.

Figure 5.—Run 4 results.

(a) RMS level.

(b) FM4.

(c) FMO.

(d) NA4.
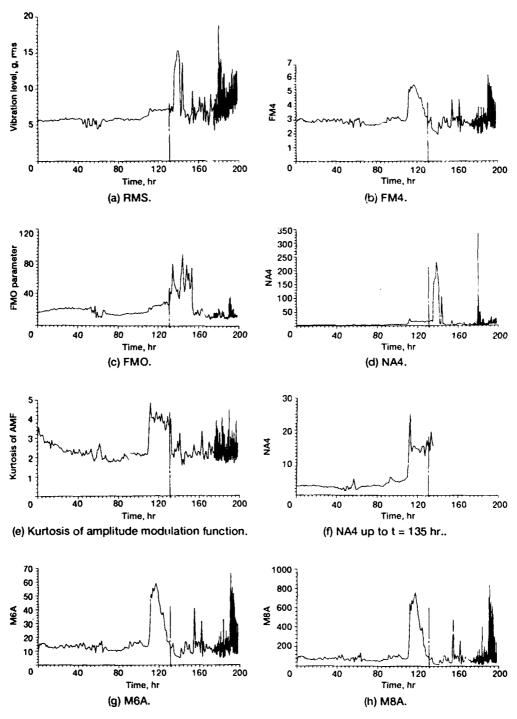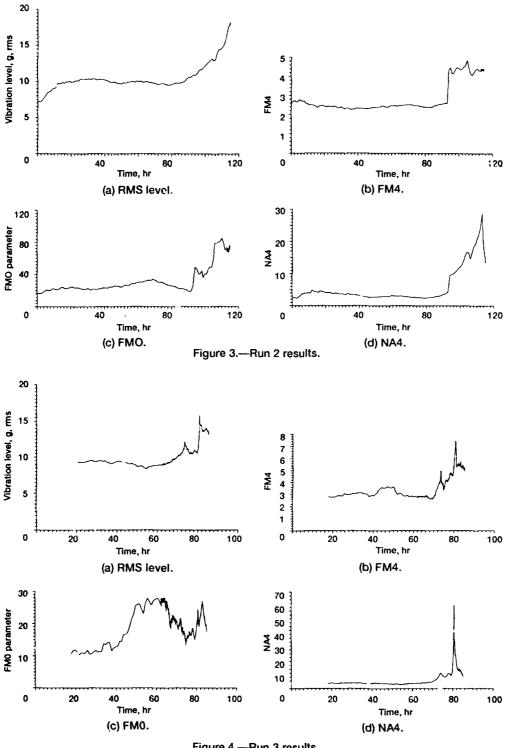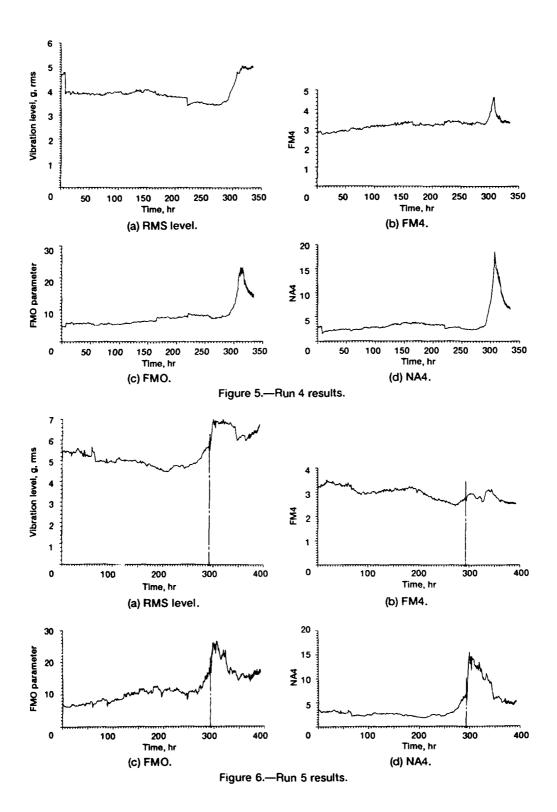
Figure 6.—Run 5 results.

SHIPBOARD OIL ANALYSIS
A PROACTIVE MAINTENANCE APPROACH

Commander G.R. Baker Royal Navy
Carderock Division
Naval Surface Warfare Center
Philadelphia, PA 19112

ABSTRACT: The U.S. Navy, like the Royal Navy, is to a large part playing lip service to the requirements of a comprehensive fluid hygiene program for lubricants on board its ships and submarines. Experience over the past several years, specifically with hydraulic systems, has shown that even minute particle contamination, as small as 5 microns, can have a significant detrimental effect on lubricated mechanical systems. The life span and reliability of our machines, engines and systems, which depend on adequate lubrication, can be dramatically improved by a simple concept called proactive maintenance. Proactive maintenance is a concept that identifies the root cause of equipment degradation and strives to correct the same before mechanical wear is initiated. Conditions are maintained that avoid the onset of machine wear and component failure. This concept will extend the life of mechanical equipment and systems, minimize untimely breakdowns and reduce the need and budget allocations for unseen emergency repairs. The environmental impact of the Navy's lubrication methodology is also significant. Because periodic lubricating oil changes are inherently conservative, lubricants with remaining useful life are routinely discarded. A comprehensive shipboard oil analysis program will enable the shipboard technician to accurately determine the serviceability of lubricants and schedule oil change outs based on actual lubricant condition; thus eliminating the need for costly time based lubricant replacements.

KEY WORDS: Condition based maintenance; contamination; diesels; filters; gas turbines; hydraulics; lubrication; maintenance; oil; proactive; savings; ventilation.

INTRODUCTION: Mechanical systems have long needed lubricating oils for their successful operation. Ever since man invented the wheel, there has been a requirement for lubrication to overcome the forces of friction. Not only does lubrication significantly reduce the wear between two rubbing surfaces, but early man soon learned that the amount of effort required in pulling his cart is greatly reduced if the cart's wheel bearings were adequately lubricated. If the cart became too hard to pull, more "Grease" was added. This may have been the first indications of proactive maintenance. Unbeknown to early man, by making the cart easier to pull, he was also reducing the wear on the sliding components and making his cart last longer.

There is a great deal of analogy that can be made with the human body and the blood system. The blood system is designed to carry nutrients to various components throughout the body and to remove waste products. Those waste products are filtered out and disposed of separately while the

clean blood continues on another mission around the body. When we are ill, doctors first test the blood to see if anything can be determined. It is the same with oil. Oil is needed to remove the waste products of machinery operation. It removes unnecessary heat and carries away contaminants. Heat is removed by exchanges (Oil Coolers) and contamination by filters. By measuring the health of the equipment's oil, we can determine very accurately the health of that equipment.

However, unlike our medical counterparts, it is only recently that oil has been used to help determine the causes of failure. Wear Debris Analysis, Spectrographic Analysis and Ferrography all rely on wear taking place. These techniques then look at the resultant wear in an attempt to discover what is failing, how long the component will last and when maintenance needs to be scheduled. These are the current cornerstones of condition based, predictive maintenance systems. While very good in their own right, these techniques do nothing to extend the life of the machine/component in the first place? The good engineer does not want to know when his machine is going to fail, but how can he prevent the machine from failing in the first place. How can he make his machinery last longer!

Importance of Lubrication: As mentioned above, lubrication is essential to successful machinery operations. All rely totally on a successful lubricating system for their very operation. Lubricating oil has many functions:

Reduction of Friction: One of the most common and important properties of a lubricant. Friction produces unwanted heat, component wear and inefficient operation.

Heat Transfer: The oil must have a high affinity to heat so as to readily absorb excessive heat generated by friction and the operation process (i.e. Steam Turbine Journals) and carry that heat away from the bearing location. By reducing the bearing operating temperature, less heat sensitive bearing materials may be used.

Contaminants: The oil must be capable of keeping itself clean with the aid of good filtration. Contamination may be present in the oil from a variety of sources including the atmosphere, new oil makeup, debris from construction and maintenance processes as well as wear debris from the operation of the equipment. The oil must carry these contaminants away to the filtration equipment where a good, efficient and effective filter can remove them successfully.

Control: Many equipments utilize lubricating oil as a control medium. The mechanical control systems that use lubricating oil typically have fine clearances. The lubricating fluid must therefore be clean if unstable conditions or failures are to be avoided.

Protection: The lubricating oil must protect against wear and against corrosion. Degradation of component internal surfaces will result in loss of material, structural strength and, more importantly, will increase the level of contaminants in the fluid.

Vital U.S. Navy Equipment Oil Lubricated: A warship is a complicated combination of complex equipment and systems from the simple motor boat engine, the gas turbine and gearbox through to the weapon system launchers and other items necessary for the defense of the homeland. These equipments contain a lubrication system which has been managed in a similar manner for many years. Ship's force are instructed in Naval Ship Technical Manual (NSTM) Chapter 262 how and when to test the oil. Typically, the tests are viscosity, acid number and visual.

NOAP Program: The Naval Oil Analysis Program was designed to provide ships force and maintenance personnel with the facilities of an oil monitoring laboratory to supplement the shipboard oil monitoring program. It applies to almost every piece of oil lubricated machinery on the ship. Samples are drawn by ship's force and sent to the nearest NOAP laboratory for analysis. Typically, for diesel engine lubricating oil, the tests will include spectroscopy, fuel dilution, viscosity, acidity (Neutralization Number) and water content. For hydraulic oils, tests include particulate counts and water content. While successful, the program requires that each diesel engine's oil be sampled and tested every 100 hours of operation. On a diesel powered ship this equates to one sample per diesel engine every 4 days; all of which are dispatched to the nearest NOAP Laboratory. Most results are obtained in several days, but delays of several weeks are not unknown. Reports are in message format with recommendations including satisfactory, re-sample and change the oil. Little advice is offered as to machine health, nor trending for NOAP results. The trending problem is further compounded by the routine change out of oil on time rather than condition.

MAINTENANCE SYSTEMS:

Breakdown Maintenance: This is the simplest to operate. Run the equipment and wait for it to break down.

Preventive Maintenance: Particular pieces of equipment are tested to determine their life expectancy and maintenance is scheduled before failure is anticipated to occur. It takes no account of equipment condition, nor the actual need for maintenance.

Predictive Maintenance: The equipment is monitored during its operation and only when some parameter that indicates machine degradation has occurred, will maintenance be scheduled. This degradation may give rise to increases in wear metals and vibration signatures, flow losses, higher energy consumption or abnormal temperatures. Nothing in predictive maintenance will extend the life of the machine. The program is designed

211

to allow the operator to program the necessary down time to correct the problem before a more serious fault occurs or the machine fails completely.

Proactive Maintenance: This is the methodology which monitors the operation of machinery from the initial correct installation and looks for the root causes of machine degradation.   For machine wear and failure to occur, some parameter must be outside the original design specification. These parameters may include temperature, load, speed, contamination, chemical, pressure and environmental.   The objective of proactive maintenance is to monitor those critical parameters which give rise to machinery degradation and wear and to correct those problems BEFORE the machinery begins to degrade.

PROACTIVE MAINTENANCE: Proactive maintenance is not a new concept.   As the strategy unfolds, it will become clear that some of the maintenance practices currently done under the mantle of condition based maintenance are in fact proactive.   Proactive maintenance has one underlying theme: monitor the operation of machinery, understand the root causes of machinery degradation and failure, monitor those root causes and, when one or more of those root causes change in such a way as to cause the machine distress, correct the root cause aberration before the machine degrades. It is the implementation of a strategy to monitor those root causes of failure that yields significant savings in plant and equipment maintenance budgets. The strategy must apply to a complete piece of equipment, system or overall plant.   One must monitor the root causes of machinery degradation and, when an abnormal condition exists, correct the root cause before machinery degradation occurs. Because no condition is stable, the maintenance program must also monitor the effects of the remedial actions to ensure successful removal of the root cause of failure.   Proactive maintenance has three phases. All three are important and the successful implementation of one is essential to the success of the others.

Installation Phase: The key to good, effective machinery operation is a sound initial installation. Efforts up front always pay dividends in the long term. Modern techniques, including vibration monitoring, should be used to ensure that the installation is correct.   Using vibration monitoring to ensure that the equipment is correctly aligned is proactive maintenance as a root cause of failure - misalignment - is removed.   If left uncorrected, the misalignment may cause bearing damage.   It is insufficient to monitor the bearing for damage; the root cause of failure, misalignment, must be corrected.   Also, if the balance is not correct, bearing wear may occur.   The root cause in this case is imbalance. Proactive maintenance on installation will use vibration monitoring as a tool to identify root causes of machinery failure which are active and allow for their correction before the equipment is put to general use.

Operational Phase: Once machinery has been commissioned and is running as designed, the true role of proactive maintenance comes into force. Root causes of machine degradation are monitored regularly so that when one is detected out of parameter, action can be planned to correct that

212

deficiency before the machine degrades. The following are two typical examples but the list is only limited by the imagination and creativity of the maintenance engineer and his staff.

Diesel Engines: The successful operation of diesel engines relies heavily on the fuel injection system. The correct operation of fuel injectors for timing, spray pattern, cut-off and operating pressure relate directly to the efficient operation of the engine as a whole. A poor injector will reduce performance, increase smoke production and, most importantly, introduce unburnt fuel into the cylinders. That fuel will condense on the cylinder liners and remove the protective layer of lubricant. If allowed to persist, liner and ring wear will increase, the engine will lose compression and, ultimately, fail. A major overhaul will then be required. One simple action for proactive maintenance would be to observe the engine's exhaust gases. If a smokey exhaust is evident, check the fuel system and correct any root cause, in this case a mal-adjusted injector. A more significant action would be to regularly test the lubricating oil for fuel dilution. Any significant change over a short period of time will indicate a problem. The corrective action would be to check the fuel injection system, change out or reset incorrectly set/failed injectors as required and return the engine to service. By catching the problem early, not only is the engine protected from further wear, but, as shown in Figure 1, the main lubricating oil is only at 0.5% fuel dilution, still good for many more hours of service.
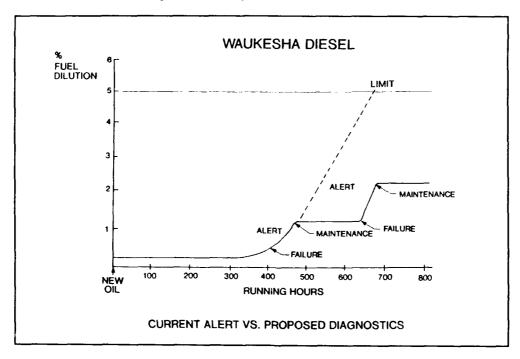
Figure 1. Diesel Engine Life Extensions

In addition, further tests, such as viscosity, TBN and particulate contamination, are available to monitor engine condition. Acidity and neutralization number are no longer meaningful with today's TBN package in diesel oil. Any acid formed by the combustion process is neutralized by the base additive. Therefore it is necessary to monitor the base additive package for depletion. Any depletion greater than the normal rate would indicate a root cause problem and should be investigated. The same can be applied to viscosity. Particulate contamination will be dealt with later in the paper.

By the adoption of the above monitoring programs, it can be seen that Lubricating Oil is an important element in the proactive maintenance strategy. Regular testing will identify root causes out of limits and enable equipment operators to change out the lubricating oil on condition only. Currently, FFG 7 Ship Service Diesel Generator lubricating oil change outs are required every year or 2000 hours, whichever occurs first. For continuous operation, 2000 hours equates to three months. A recent oil test of a Royal Navy Diesel Generator which had operated for 8000 hours without overhaul, and with an oil charge not replaced within the last 18 months showed a fuel dilution of 0.5%, viscosity 125 cSt and a TBN of 7.8. Not only did this indicate a well running engine, but a lubricating oil with much life left.

If a problem is suspected to exist and root cause correction has not solved the problem, then the maintainers can call for a NOAP analysis. This may point to an area of concern. However, NOAP analyses, as well as wear debris and ferrography, rely on machinery degradation to have already occurred, otherwise there would be no material to analyze. The techniques will not prolong the life of the equipment, only mitigate the wear that has already occurred.

Ventilation Equipment: Ventilation equipment is often one of those equipments which operates unseen until something goes wrong. In a recent case, ventilation fans serving a foundry were suffering a series of bearing failures. The bearings were monitored and, when wear was detected, maintenance actions were scheduled to minimize down time. The failure rate of the bearings continued to be the same. On investigation, the root cause of the bearing failures determined to be imbalance in the ventilation fan's impeller. The root cause of this imbalance was a build up of foundry grime. The foundry's maintenance personnel installed vibration monitoring equipment set to measure imbalance. Once a preset alarm level was reached, the fan was scheduled for a cleaning. The alarm level was set below the level likely to cause bearing damage. The root cause, imbalance, was corrected before machine degradation occurred. By monitoring the balance both before and after the repair, a feedback loop was established. The bearing replacement requirements were significantly reduced, saving material costs and equipment down time. This is truly proactive maintenance.

Repair/Replacement: Despite all the effort expended by maintenance personnel, equipment failures do occur. Accordingly, repairs and replacements must be accomplished. Proactive maintenance plays a key role in this stage of machinery life. Examination of failed parts is made to

214

identify the actual cause of failure and to determine if adjustments to
the equipments's maintenance is warranted. Post repair/replacement
testing is also conducted as described in the installation phase to ensure
that infant mortality problems are minimized.

Equipment Life Expectancy: Figure 2 shows the typical curve used by
maintenance engineers to show the operational and wear out phases of a
machine's life. Also shown is the area affected by the deficient
maintenance strategies explained above. Breakdown maintenance has no
effect on the operational life of the equipment. By the time breakdown
occurs, the only solution is total equipment repair. The predictive
maintenance range requires some form of degradation to occur before the
effects can be monitored. Monitoring by wear debris analysis, vibration
analysis, infrared thermography, etc. is only effective once wear is
initiated. Hence machine life cannot be extended. The inevitable machine
failure can only be delayed and, at best, avoided by scheduling
maintenance. Only proactive maintenance has the benefit of extending
machine life. By removing the root cause of equipment wear and failure,
wear is minimized. The whole of the curve can then be moved to the right.
With careful management, the ultimate wear out is a factor of the original
design and not one of operation.
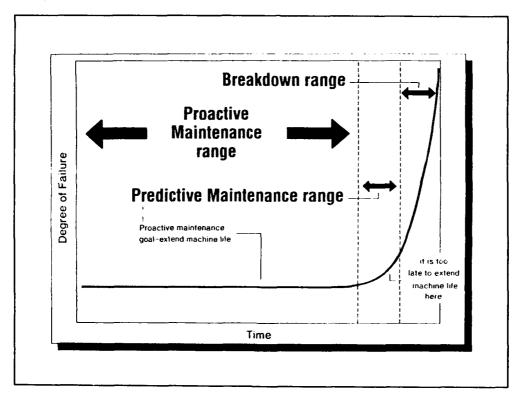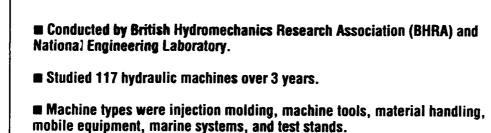
Figure 2. Benefits of Proactive Maintenance

215

LUBRICATING OIL: Lubricating oil is the hidden blood of our machines. As explained above, it is the very essence of successful machine operation. Neglect lubricating oil and machinery will fail. A comprehensive oil management program, from initial fill, maintenance and disposal is essential to any engineering operation. Simple steps such as sealed oil containers, contents markings and sealed storage tanks with replenishing lines will go a long way to prevent contamination and to ensure that the correct oil is placed in the correct machine. These measures are obvious. The hidden enemy is **CONTAMINATION.** Among the preventable cause of machine failure, contamination is number one. The old adage that "Clear and Bright" indicates everything is fine is not good enough. The human eye, at best, can see only 40 microns and larger. Studies have shown that contamination as small as 5 microns will damage machinery. The contaminant will imbed itself into the system's softer material and ultimately flake off a piece of harder material. This will increase the level of contamination in the system and, unless action is taken promptly, machine wear rates will increase until machine failure results.

Studies by the British Hydromechanical Research Association (BHRA) have demonstrated that the theory proposed in laboratory tests have a direct applicability in the field. Figure 3 amply demonstrates that 10 to 50 times life extensions for hydraulic equipment can be achieved by improving

■ **Conducted by British Hydromechanics Research Association (BHRA) and National Engineering Laboratory.**

■ **Studied 117 hydraulic machines over 3 years.**

■ **Machine types were injection molding, machine tools, material handling, mobile equipment, marine systems, and test stands.**
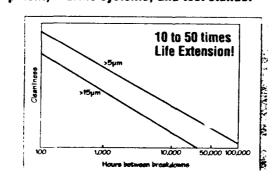
Figure 3. BHRA Machine Failure Study

cleanliness levels. Similarly, Figure 4 demonstrates how NIPPON Steel, one the world's largest steel producers, showed a dramatic improvement in plant operation and reduction in pump replacement once a proactive maintenance program was introduced. Contamination is the invisible source of machine failure. If contamination is controlled up front and viscosity, lubricity and additives are maintained, then the major battle in the war against machinery wear and failure has been won.
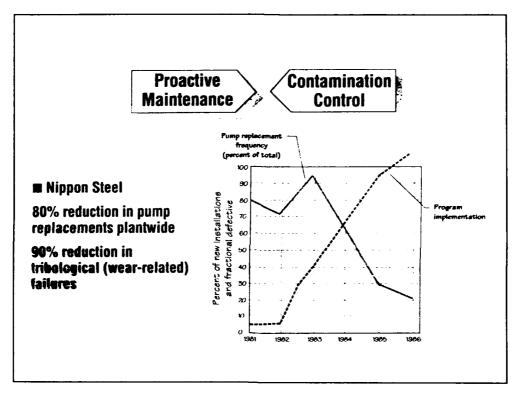
Figure 4. Nippon Steel Contamination Control Study

Why Monitor Contamination: It is important to know the contamination level in each equipment's lubricating oil. Only through periodic monitoring can the levels be accurately known and trended. This is the first step to proactive maintenance. High levels of contamination are a root cause of machine wear and failure. Monitor the levels, keep them below a well defined limit below which machinery wear is greatly reduced and the life of the equipment will be prolonged. It is essential to monitor the contamination levels of various parts of the system. The efficiency of filters is easily established by measuring the contamination both above and below the filter. This has a two fold advantage, especially for a return line filter which would give early indication of

a potential problem. The frequency of filter change out is established by true condition and not some arbitrary time interval. Additionally, the traditional method of differential pressure gauges is unreliable and unsafe. Experiments have determined that filters de-absorb before the pressure differential increase gives cause for concern. Once a degradation in the contaminant level is detected, the root cause of that contamination must be established and eliminated and the system filtered clean again before resuming safe operation. This action will prevent the machine degradation from occurring, correct the root cause and prolong machine life.

TESTING TECHNIQUES: The Naval Ship Systems Engineering Station has for some time been designing, testing and installing revised lubricating oil equipment for the U.S. Navy. The mainstays of shipboard test equipment are the Drop Ball Comparator for viscosity/fuel dilution, the neutralization number for acidity in diesel oils, and water. The viscosity of diesel engine oil can be altered by two factors. The viscosity will increase with usage, lacquer deposits, oxidation and contamination while fuel dilution will decrease the viscosity. It is not improbable that an oil may be unfit for service, say with >5% fuel dilution but because of lacquering and oxidation, the overall viscosity is within range. The test for viscosity still remains very subjective, relying on the ability and experience of the operator, and acts as a go/no go gauge only. The test equipment being evaluated by NAVSSES includes a viscosity meter, fuel dilution meter, TBN meter and, under test from the commercial environment, a particulate contamination meter and water content meter. In addition to the meters, a comprehensive package of instructions and data handling is being developed by NAVSSES to ensure that this sophisticated technology is effective when used in a shipboard environment. When added to automated data collection of equipment operating parameters and vibration analysis techniques integrated into a true proactive maintenance package, the sailor will have a true picture of his equipment and systems. For the commercial user, the package offers a real solution to plant operation and maintenance, ensuring longevity of equipment, less down time and cheaper operating costs.

New NAVSSES Test Equipment:

Viscosity: In association with Cambridge Instruments, a new viscosity meter has been developed which measures absolute viscosity. This has the distinct advantage of negating the need to know the origins of the original oil and takes into account the wide procurement specifications that the U.S. Navy uses in purchasing its oil. The viscosity meter is simple to use. By measuring the time it takes for a small shuttle to move through the charge of oil when influenced by a magnetic field, the electronic circuit, allowing for the temperature of the oil, translates the signal to a direct measure of viscosity. The equipment operates in the range 100-300 cSt at 100 Deg F. The cycle time is 2 minutes and an LED readout is provided. Alarms can be set as required; currently they are set at 100 cSt LOW and 225 cSt HIGH.

TBN: With the advent of modern diesel engine oils containing a significant base additive to combat acid formation, a meter to measure the status of the base additive package was required as the current neutralization test is now meaningless. By the time that a rise in acidity levels is detected, the additive package has depleted and the engine is at risk from acid formation and attack. NAVSSES has developed a simple to use, safe and effective meter to measure the TBN levels. By using a known amount of reagent to react with the base additive in the sample, the resultant pressure rise is converted into a TBN reading . The instrument will measure 2-14 TBN, with a 10 minute cycle time.

Fuel Dilution: The fuel dilution meter measures the partial pressure of the fuel in the lubricating oil and converts it to a % dilution. It is very simple to use, has a cycle time of 3 minutes, operates in the 0-5% range, an LCD display with audio/visual alarms for warnings. In shipboard trials it has proved to be effective and easy to use. As with the viscosity meter, it requires no comparison with previous new oil.

Other Initiatives: Further work by NAVSSES includes a particulate contamination meter (Lube Oil) and water detection in fuels.

AFFORDABILITY: The program described in this paper saves money, not only in longer service life for equipment, but, as shown below, in a reduction in oil consumption costs:

Example: FFG 7 Ships Service Diesel Generator:

Oil Charge: 245 gallons of Mil Spec 9250 @ $2.80 per gallon	$686
Cost to dispose of used 245 Gallons at $20 per Gallon (PH)	$4,900
Time between oil changes Annual or 2000 hours (3 months running)	
2 year cost per ship (2 oil changes/year for 4 diesels)	$89,376
Cost to change oil on condition only (One change every 2 years per RN experience)	$22,344
Savings per ship:	$67,032

Cost of test equipment for one ship set:

TBN Meter	$2500
Fuel Dilution Meter	$3500
Viscosity Meter	$5000
Contamination Meter	$9000

	$20,000
Net savings per ship:	$47,032
Total savings for 51 FFG 7s in **2 Years:**	**$2,398,632**

In addition, there will be less new oil to buy and, more importantly for the environment, much less used oil to be disposed.

PROGRAM IMPLEMENTATION: The main thrust of program implementation would be to introduce the shipboard program described above to the fleet as a MACHALT with the test equipment brought into service as GPETE - General Purpose Electronic Test Equipment. NAVSSES has the ability to procure the initial outfits of equipment, the resources to develop the necessary training packages and the personnel required to install the shipboard system fleetwide. Integration with an automatic watch keeping information system downloaded to combine with information normally gathered by watch standers on paper is planned. This is achieved by developing and commissioning a 486 based data capture system to log and trend the very watch keeping information that ships produce every hour of every day. Also added is a vibration monitoring and analysis package. When combined with watch keeping information, accurate information is available to assess the genuine condition of shipboard equipment. By adding the valuable information from good oil analysis from the onboard test equipment, supplemented by shore NOAP analysis where necessary, all the ingredients are present to monitor for root causes of mechanical degradation. By highlighting these shortcomings, ship's force is in an ideal position to correct root causes before mechanical degradation occurs and thus prolong the service life of shipboard equipment.

CONCLUSIONS: Significant savings can be made by controlling the level of particulate contamination present in oil systems as a two fold increase in life can be achieved by a modest increase in cleanliness levels. By setting well engineered contaminant levels for all in service equipment, setting the correct standards and providing modern, up to date test equipment, even more savings can be realized. These savings are represented by less equipment down time, less spares replacement, less overhaul requirements and less oil consumption and disposal requirements. All that is needed is the will to drive the program forward.

Lubricating oil is essential to the operation of a wide range of machinery. While no one will argue this point, the U.S. Navy, to a large degree, tends to pay lip service to the requirements for a comprehensive fluid hygiene program for lubricants aboard its ships and submarines. Limited maintenance budgets and the costs associated with replacing lubricating oils (both financial and environmental) mandate that the Navy's maintenance managers take steps to tap the potential of lubricating oil diagnostics as a tool of proactive maintenance and to stop the unnecessary change out of good oil. The principles and equipment required to take these steps exist and are not difficult to comprehend. It is up to us to embrace them and to strive forward.

220

PATTERN CLASSIFIER FOR HEALTH MONITORING OF HELICOPTER GEARBOXES[1]

Hsinyung Chin, Graduate Research Assistant
Kourosh Danai, Assistant Professor
Department of Mechanical Engineering
University of Massachusetts
Amherst, MA 01003

and

David G. Lewicki
NASA Lewis Research Center
Cleveland, OH 44135

Abstract: The application of a newly developed diagnostic method to a helicopter gearbox is demonstrated. This method is a pattern classifier which uses a *multi-valued influence matrix (MVIM)* as its diagnostic model. The method benefits from a fast learning algorithm, based on error feedback, that enables it to estimate gearbox health from a small set of measurement-fault data. The MVIM method can also assess the diagnosability of the system and variability of the fault signatures as the basis to improve fault signatures. This method was tested on vibration signals reflecting various faults in an OH-58A main rotor transmission gearbox. The vibration signals were then digitized and processed by a vibration signal analyzer to enhance and extract various features of the vibration data. The parameters obtained from this analyzer were utilized to train and test the performance of the MVIM method in both detection and diagnosis. The results indicate that the MVIM method provided excellent detection results when the full range of faults effects on the measurements were included in training, and it had a correct diagnostic rate of 95% when the faults were included in training.

Key Words: Detection; diagnosis; helicopter gearbox; pattern classification; vibration signal processing

Introduction: Helicopter drive trains are significant contributors to both maintenance cost and flight safety incidents. Drive trains comprise almost 30% of maintenance costs and 16% of mechanically related malfunctions that often result in the loss of aircraft [6]. As such, it is crucial that faults be detected and diagnosed in-flight so as to prevent loss of lives.

Fault diagnosis of helicopter gearboxes is based primarily on vibration monitoring and extraction of features that relate to individual gearbox components. Therefore, considerable effort has been directed toward the development of signal processing techniques which can quantify such features through the parameters they estimate (e.g., [13,15]). For example, the crest factor of vibration, which represents

---

[1]This paper is extracted from References [4] and [5].

the peak-to-rms ratio of vibration, has been shown to increase with localized faults such as tooth cracks [1]. However, due to the complexity of helicopter gearboxes and the interaction between their various components, the individual parameters estimated from vibration measurements do not provide a reliable basis for fault detection and diagnosis.

As an alternative to single-parameter based diagnosis, fault signatures can be established so as to consist of many parameters. For this purpose, pattern classification techniques need to be employed [9,14]. Among the various pattern classifiers used for diagnosis, artificial neural nets are the most notable due to their nonparametric nature (i.e., independence of the probabilistic structure of the system), and their ability to generate complex decision regions [16]. However, neural nets generally require extensive training to develop the decision regions. In cases such as helicopter gearboxes, where adequate data may not available for training, neural nets may produce false alarms, undetected faults, and/or misdiagnoses.

In this paper we demonstrate the application of a diagnostic method that can estimate gearbox health based on a small set of measured vibration data. This method uses nonparametric pattern classification in its model, so like artificial neural nets, is independent of the probabilistic structure of the system. Moreover, it utilizes a *multi-valued influence matrix (MVIM)* as its diagnostic model that provides indices for diagnosability of the process and variability of the fault signatures [8]. These indices are used as feedback to improve fault signatures through adaptation [7].

To test this method, vibration signals were collected at NASA Lewis Research Center as part of a joint NASA/Navy/Army Advanced Lubricants Program to reflect the effect of various faults in an OH-58A main rotor transmission gearbox. In order to identify the effect of faults on the vibration data, the vibration signals obtained from five tests were digitized and processed by a vibration signal analyzer. The parameters obtained from this signal analyzer were then utilized to train the MVIM method and test its performance in both detection and diagnosis.

MVIM Method: Measurements are processed in the MVIM method as illustrated in Fig. 1: They are usually pre-processed first to obtain a vector of processed measurements **P**, then they are converted to binary numbers through a flagging operation (i.e., abnormal measurements characterized by 1 and normal ones by 0) to obtain a vector of flagged measurements **Y**, and finally they are analyzed through the diagnostic model to produce fault vector $\hat{\mathbf{X}}$. The MVIM method is explained in detail in [3] and [7], and its overall concept is briefly discussed here for completeness.

Fault Signature Representation: Fault signatures in the MVIM method are represented by the n unit-length columns $\mathbf{V}_j \in \mathcal{R}^m$ of a multi-valued influence matrix (MVIM) **A**:

$$\mathbf{A} = [\mathbf{V}_1 \ \ldots \ \mathbf{V}_j \ \ldots \ \mathbf{V}_n] \tag{1}$$

where m denotes the number of characteristic parameters processed from the raw
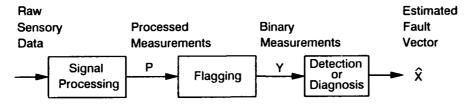
Figure 1: Processing of measurements in the MVIM method.

data, and n represents the number of different fault conditions, including the no-fault condition.

Diagnostic Reasoning: In the MVIM method, the fault vector $\hat{X}$ which ranks the faults according to their possibility of occurrence is defined by the closeness of the influence vector to the vector of flagged measurements Y (see Fig. 2).
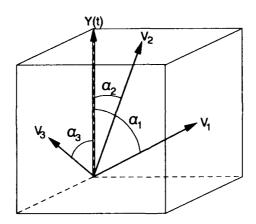

Figure 2: Schematic of diagnostic reasoning in the MVIM method, illustrated in three dimensional space.

Fault Signature Evaluation: The influence vectors defined in Eq. (1) are not known *a-priori* and need to be estimated. In the MVIM method, the error in diagnosis is used as the basis to estimate/update the influence vectors. For this purpose, the fault signatures are updated recursively after the occurrence of each fault to minimize the sum of the squared diagnostic error associated with that fault [8].

One of the unique features of the MVIM method is its ability to evaluate quantitatively the uniqueness of the fault signatures as well as their variability, so that these quantitative measures can be used to improve the flagging operation. In the MVIM method, the uniqueness of fault signatures is characterized by the closeness of pairs of influence vectors. For this purpose, a diagnosability matrix is defined

to represent the closeness of the orientation of individual influence vectors [8], and the index of diagnosability is defined as the smallest off-diagonal component of this matrix so as to denote the closest pair of fault signatures.

In the MVIM method, the variability of fault signatures is defined by their variance. For this purpose, the variance matrix associated with $\bar{A}$ is estimated to provide a measure of the variations of individual components of the influence matrix. Since in the MVIM method the components of $\bar{A}$ are adjusted recursively, the variance matrix can be readily estimated during training [7]. The index of fault signature variability in the MVIM method is defined as the largest component of a variance matrix which represents the variability in the components of matrix $\bar{A}$.

Flagging Unit: The influence matrix $\bar{A}$ is estimated based on the values of the flagged measurement vector Y. Thus, before the influence matrix is used for diagnostic reasoning, the integrity of the flagging operation needs to be ensured. Ideally, the measurements should be flagged such that no false alarms are produced, all faults are detected, the fault signatures are as spread out as possible, and the variability of flagged measurements for individual faults is minimized. To this end, a Flagging Unit is designed so that it can be tuned to achieve the above goals. The Flagging Unit is tuned iteratively based on a training batch, where at the end of each iteration the total number of false alarms and undetected faults are counted and the uniqueness and variability of the fault signatures are obtained from MVIM. This information is then used as feedback in the next iteration to improve the performance of the Flagging Unit (see Fig. 3). Training stops when the total number of false alarms and undetected faults are minimized, and the uniqueness and repeatability of fault signatures are enhanced [7].
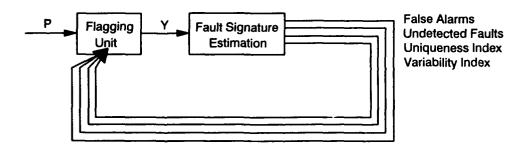
Figure 3: Iterative tuning of the Flagging Unit based on feedback from its diagnostic model.

Experimental: Vibration data was collected at NASA Lewis Research Center to reflect the effect of various faults in an OH-58A main rotor transmission gearbox [11]. The gearbox was tested in the NASA Lewis 500-hp helicopter transmission test stand providing an input torque level of about 3100 in-lbs and an input speed of 6060 rpm. The configuration of the gearbox is shown in Fig. 4. The vibration signals were measured by eight piezoelectric accelerometers (frequency

range of up to 10 kHz), and an FM tape recorder was used to record the signals periodically once every hour, for about one to two minutes per recording (at the tape speed of 30 in/sec, providing a bandwidth of 20 kHz). Two chip detectors were also mounted inside the gearbox to detect the debris caused by component failures. The location and orientation of the accelerometers are shown in Fig. 5.
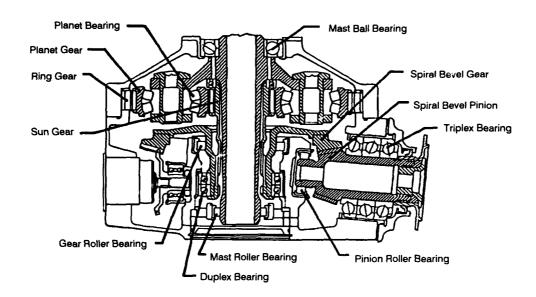
Figure 4: Configuration of the OH-58A main rotor transmission gearbox.

During the experiments, the gearbox was disassembled/checked periodically or when one of the chip detectors indicated a failure. A total of five tests were performed, where each test was run between nine to fifteen days for approximately four to eight hours a day. Among the eight failures which occurred during these tests, there were three cases of planet bearing failure, three cases of sun gear failure, two cases of top housing cover crack, and one case each of spiral bevel pinion, mast bearing, and planet gear failure (see Table 1). Insofar as fault detection during these tests, the chip detectors were reliable in detecting failures in which a significant amount of debris was generated, such as the planet bearing failures and one sun gear failure. The remaining failures were detected during routine disassembly and inspection.

Signal Processing: In order to identify the effect of faults on the vibration data, the vibration signals obtained from the five tests were digitized and processed by a commercially available signal analyzer [17]. For analysis purposes, only one data record per day was used for each test. These data records were taken at the beginning of the day unless a fault was reported, which in that case, the record taken right before the fault incident was selected to ensure that the data record

225
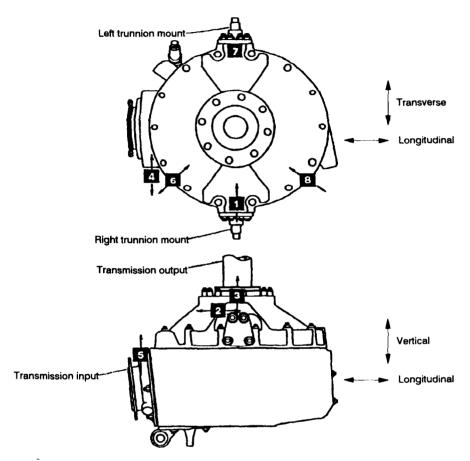
Figure 5: Location of the accelerometers on the test stand.

reflected the fault. Also, in order to reduce estimation errors, each data record was partitioned into sixteen segments and parameters were estimated for each segment and averaged over these segments. A total of fifty-four parameters were obtained, of which nineteen parameters were obtained for statistical analysis, baseband power spectrum analysis, and bearing analysis. The other thirty-five parameters reflected the various features of signal averaged data (seven parameters for each of the five gears) [2].

Implementation: As explained earlier, the MVIM method requires a set of measurements during normal operation and at fault incidents to estimate the no-fault and fault signatures. The parameters obtained from the signal analyzer were utilized to evaluate the performance of the MVIM method, first in detection and then in diagnosis.

226

Test #	Number of Days	Failures
1	9	Sun gear tooth spall Spiral bevel pinion scoring/heavy wear
2	9	None
3	13	Planet bearing inner race spall Top cover housing crack Planet bearing inner race spall Micropitting on mast bearing
4	15	Planet bearing inner race spall Sun gear tooth pit
5	11	Sun gear teeth spalls Planet gear tooth spall Top housing cover crack

Table 1: Faults occurred during the experiments.

Fault Detection: The mean values of the nineteen "non-signal averaged" parameters were used as the components of the measurement vector **P** (see Fig. 1) to train and test the MVIM method in detection. Since signal averaging is usually time consuming and may not be suitable for on-line detection [12], the thirty-five "signal averaged" parameters were not utilized for detection. For scaling purposes, each parameter value was normalized with respect to the value of the parameter on the first day of each test. Since in the experiments the exact time of fault was not known, the exact times for the fault incidents of the five tests needed to be established before the measurements could be used for training and testing the MVIM. For this purpose, *Kohonen's feature mapping* [10], an unsupervised learning algorithm, was first used to classify individual parameters into no-fault and fault cases. The exact time of fault incidents was then established through correlating these parameters with the faults which had been detected in each test [2]. The status of various faults during the five tests are shown in Table 2.

The effectiveness of the MVIM method in detection was evaluated with various training sets. For this purpose, training sets were formed based on parameters from various combinations of the five tests (see Table 3). The MVIM was tested, however, based on the parameters from all of the five tests. For each training case, the MVIM was iteratively trained until perfect detection was achieved within the training set (i.e., no false alarm or undetected fault was found in the training set). Note that the MVIM trained for detection contains only two columns, one representing the no-fault signature and the other representing the fault signature. The detection results produced by the MVIM for eighteen different cases of training are shown in Table 3. For comparison, the results obtained from the MVIM method are contrasted against the results obtained from a multilayer neural net which was trained and tested under the same conditions. Performance of these detection

Day	Fault Status				
	Test #1	Test #2	Test #3	Test #4	Test #5
1	x_0	x_0	x_0	x_0	x_0
2	x_0	x_0	x_0	x_0	x_0
3	x_0	x_0	x_2	x_0	x_0
4	x_0	x_0	x_2	x_0	x_0
5	x_4	x_0	x_0	x_0	x_0
6	x_4	x_0	x_0	x_0	x_0
7	x_4	x_0	x_0	x_0	x_3
8	x_4	x_0	x_0	x_0	x_3
9	x_4, x_1	x_0	x_3	x_0	x_3
10			x_0	x_0	x_3, x_1
11			x_2	x_2	x_3, x_1, x_5
12			x_2	x_2	
13			x_6	x_0	
14				x_1	
15				x_1	

Table 2: Association of data from each day of the five tests with no-fault and various fault cases. The no-fault case is denoted as x_0 and the six faults are represented as x_1: sun gear failure, x_2: planet bearing failure, x_3: housing crack, x_4: spiral bevel pinion failure, x_5: planet gear failure, x_6: mast bearing failure.

methods are represented by the total number of false alarms and undetected faults they produced during testing (denoted as "Total Test Errors" in Table 3).

The results in Table 3 indicate that the MVIM was able to provide perfect detection when faults were fully represented by the training sets (i.e., Cases #10, #11, #13, #16, #17, and #18), and that it produced better results than the Net in most of the cases. Specifically, the MVIM produced better results in twelve of the test cases, produced identical results in five cases, and was outperformed in only one case. Upon a casual inspection of the training sets that enabled MVIM to perform perfect detection, it can be observed that Tests #3 and #4 were included in all of them. This implies that the MVIM needed the parameters from these two tests to establish an effective pair of signatures for no-fault and fault cases. Note that without Test #3, the MVIM produced one undetected fault and one false alarm (Case #15), and without Test #4 it produced one undetected fault (Case #14). Note that the Net could not provide perfect detection even when trained with all of the five tests (Case #18).

Case #	Training Data Sets	Diagnostic Method	Undetected Faults	False Alarms	Total Test Errors
1	1	Net	4	0	4
		MVIM	1	3	4
2	5	Net	1	2	3
		MVIM	3	2	5
3	1,2	Net	4	0	4
		MVIM	2	2	4
4	1,3	Net	1	2	3
		MVIM	2	0	2
5	2,5	Net	3	2	5
		MVIM	3	2	5
6	3,4	Net	2	2	4
		MVIM	0	0	0
7	3,5	Net	0	3	3
		MVIM	1	0	1
8	4,5	Net	3	0	3
		MVIM	1	1	2
9	1,2,5	Net	1	2	3
		MVIM	1	2	3
10	1,3,4	Net	1	0	1
		MVIM	0	0	0
11	2,3,4	Net	2	0	2
		MVIM	0	0	0
12	2,3,5	Net	1	2	3
		MVIM	1	0	1
13	1,2,3,4	Net	2	0	2
		MVIM	0	0	0
14	1,2,3,5	Net	2	1	3
		MVIM	1	0	1
15	1,2,4,5	Net	1	1	2
		MVIM	1	1	2
16	1,3,4,5	Net	1	0	1
		MVIM	0	0	0
17	2,3,4,5	Net	2	0	2
		MVIM	0	0	0
18	1,2,3,4,5	Net	1	0	1
		MVIM	0	0	0

Table 3: Detection results obtained from MVIM and a multilayer neural net when trained with different data sets.

229

<u>Fault Diagnosis:</u> All of the fifty-four parameters obtained from the signal analyzer were used to train and test the MVIM in diagnosis. The configuration of the MVIM as applied to fault diagnosis of the OH-58A gearbox is illustrated in Fig. 6. As shown in this figure, two MVIMs were used for each accelerometer. One MVIM to perform detection (i.e., to determine whether a fault had occurred or not), and a diagnostic MVIM to isolate the fault. The detection MVIM contained only two columns to characterize the no-fault and fault signatures, whereas the diagnostic MVIM contained seven columns, one characterizing the no-fault signature and the other six representing the signatures of individual faults (see Table 2). Note that the two MVIMs can be perceived as filters with different resolutions. Test #3 and #4 contained most of the failure modes (i.e., four out of six). Therefore, the parameters from these two tests were used to train the MVIMs. Note that not all of the failure modes were included in training, so the test results were not expected to be perfect. For training the detection MVIMs, signal averaged parameters were excluded because it had already been established that the nineteen non-signal averaged parameters were adequate for detection. For training the diagnostic MVIMs, however, all of the fifty-four parameters were utilized. A maximum of fifty iterations were used for training both the detection and diagnostic MVIMs.
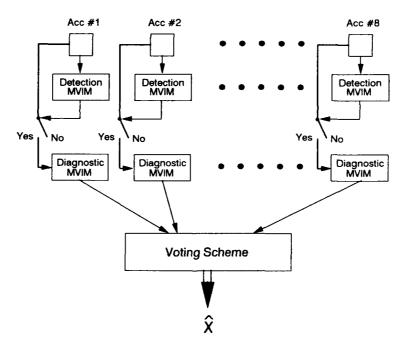
Figure 6: Configuration of the MVIM system as applied to the OH-58A main rotor transmission.

Individual MVIMs were considered converged when they produced perfect detection/diagnostics within the training set. The number of epochs for the convergence of the eight detection MVIMs were: 8, 5, 50, 37, 50, 15, 50, and 50 for accelerometers #1 to #8, respectively, whereas for the eight diagnostic MVIMs they were:

50, 1, 2, 2, 50, 50, 50, and 50. Based on the number of epochs used for individual MVIMs, it is clear that the detection MVIMs associated with accelerometers #3, #5, #7, and #8 did not achieve perfect detection within the training set. Similarly, the diagnostic MVIMs associated with accelerometers #1, #5, #6, #7, and #8 did not achieve perfect diagnosis within the training set.

The performance of the trained MVIMs were next evaluated for all of the five tests. For this purpose, the nineteen parameters from each of the eight accelerometers were first passed through the corresponding detection MVIM to reflect the occurrence of faults. Once the presence of a fault was indicated by a detection MVIM, the set of fifty-four parameters from that accelerometer was passed through the corresponding diagnostic MVIM to isolate the fault. Finally, the diagnostic results obtained from the eight diagnostic MVIMs were consolidated by a voting scheme. This voting scheme was designed based on assigning weights to individual fault signatures based on their speed of convergence in training, such that larger weights were assigned to those influence vectors which converged faster and vice versa. Zero weights were assigned to the influence vectors which did not converge during training; unity weights were assigned to those which converged within one epoch.

The diagnostic results obtained from the diagnostic system for all of the five tests are shown in Table 4, with the actual faults indicated inside parentheses. The results indicate that the MVIM system was able to produce perfect diagnostics for Tests #3 and #4, on which it was trained, and that it provided a correct diagnostic rate of 88% for all of the tests. Specifically, the results in Table 4 indicate that the MVIM system produced two false alarms (on day 4 of Test #1 and day 6 of Test #5), and five misdiagnoses (on days 5-8 of Test #1 and day 11 of Test #5). In addition, this system produced equal diagnostic certainty measures for the no-fault case (x_0) and sun gear failure (x_1) on day 10 of Test #5, and could only diagnose one of the faults on day 9 of Test #1 and on days 10 and 11 of Test #5. However, it should be noted that faults x_4 and x_5 were not included in training, so no fault signatures were estimated for them. The correct diagnostic rate of MVIM, with these two faults excluded would be over 95%, which is quite good considering that the MVIM system was trained on a small set of measurement-fault data with very few repetitions of each fault.

Summary of Results: An efficient fault detection/diagnostic system based on the MVIM method was applied to an OH-58A main rotor transmission gearbox. Detection results indicate that this system provided perfect detection when the full range of faults effects were included in training. Diagnostic results indicate that the system achieved a correct diagnostic rate of 95% despite very few repetitions of each fault.

231

Day	Estimated Fault Status									
	Test #1		Test #2		Test #3		Test #4		Test #5	
1	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)
2	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)
3	x_0	(x_0)	x_0	(x_0)	x_2	(x_2)	x_0	(x_0)	x_0	(x_0)
4	x_3	(x_0)	x_0	(x_0)	x_2	(x_2)	x_0	(x_0)	x_0	(x_0)
5	x_3	(x_4)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)
6	x_3	(x_4)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_6	(x_0)
7	x_3	(x_4)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_3	(x_3)
8	x_3	(x_4)	x_0	(x_0)	x_0	(x_0)	x_0	(x_0)	x_3	(x_3)
9	x_1	(x_4, x_1)	x_0	(x_0)	x_3	(x_3)	x_0	(x_0)	x_3	(x_3)
10					x_0	(x_0)	x_0	(x_0)	x_0, x_1	(x_3, x_1)
11					x_2	(x_2)	x_2	(x_2)	x_2, x_6	(x_3, x_1, x_5)
12					x_2	(x_2)	x_2	(x_2)		
13					x_6	(x_6)	x_0	(x_0)		
14							x_1	(x_1)		
15							x_1	(x_1)		

Table 4: Estimated faults for each day of the five tests. The actual faults (inside parenthesis) are also included for comparison. The x_i are the same as indicated in Table 2.

References

[1] Braun, S. (Ed.), *Mechanical Signature Analysis - Theory and Applications*, Academic Press, New York, NY, 1986.

[2] Chin, H., *Vibration Analysis of an OH-58A Main Rotor Transmission*, Technical Report, Department of Mechanical Engineering, University of Massachusetts, Amherst, MA, 1992.

[3] Chin, H. and K. Danai, "Improved Flagging for Pattern Classifying Diagnostic Systems," *IEEE Trans. on Systems, Man, and Cybernetics*, in press.

[4] Chin, H., K. Danai, and D. G. Lewicki, "Fault Detection of Helicopter Gearboxes Using the Multi-Valued Influence Matrix Method," *ASME J. of Mechanical Design*, in review.

[5] Chin, H., K. Danai, and D. G. Lewicki, "Efficient Fault Diagnosis of Helicopter Gearboxes," *1993 IFAC World Congress*, in review.

[6] Chin, H. and K. Danai, "Fault Diagnosis of Helicopter Power Train," *Proc. of the 1991 Annual NSF Grantees Conference in Design and Manufacturing Systems Research*, pages 787-790.

[7] Chin, H. and K. Danai, "A Method of Fault Signature Extraction for Improved Diagnosis," *ASME J. of Dynamic Systems, Measurement, and Control*, Vol. 113, No. 4, 1991, pp. 634–638.

[8] Danai, K. and H. Chin, "Fault Diagnosis with Process Uncertainty," *ASME J. of Dynamic Systems, Measurement, and Control*, Vol. 113, No. 3, 1991, pp. 339–343.

[9] Gallant, S. I., "Automated Generation of Connectionist Expert Systems for Problems Involving Noise and Redundancy," *Proc. of AAAI Workshop on Uncertainty*, 1987.

[10] Kohonen, T., *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, Germany, 1989.

[11] Lewicki, D. G., H. J. Decker, and J. T. Shimski, *Full-Scale Transmission Testing to Evaluate Advanced Lubricants*, Technical Report, NASA TM-105668, AVSCOM TR-91-C-035, NASA Lewis Research Center, Cleveland, OH, 1992.

[12] McFadden, P. D. and J. D. Smith, "A Signal Processing Technique for Detecting Local Defects in a Gear From the Signal Average of the Vibration," *Proc. of Institution of Mech. Engineers*, Vol. 199, No. C4, 1985, pp. 287–292.

[13] Mertaugh, L. J., "Evaluation of Vibration Analysis Techniques for the Detection of Gear and Bearing Faults in Helicopter Gearboxes," *Mechanical Failure Prevention Group 41th Meeting*, 1986, pp. 28–30.

[14] Pau, L. F., *Failure Diagnosis and Performance Monitoring*, Marcel Dekker, New York, NY, 1981.

[15] Pratt, J. L., "Engine and Transmission Monitoring - A Summary of Promising Approaches," *Mechanical Failure Prevention Group 41th Meeting*, 1986, pp. 229–236.

[16] Rumelhart, D. E. and J. L. McClelland (Eds.), *Parallel Distributed Processing - Explorations in the Microstructure of Cognition, Volumn 1: Foundations*, The MIT Press, Cambridge, MA, 1988.

[17] Stewart Hughes, *Transmission Systems Analysis for the MSDA, MM35: 2nd edition*, Stewart Hughes Limited, Southhampton, U. K., 1987.

PSEUDO WIGNER-VILLE DISTRIBUTION AND ITS APPLICATION TO MACHINERY CONDITION MONITORING

Young S. Shin, Jae-Jin Jeon and Scott G. Spooner

Department of Mechanical Engineering
Naval Postgraduate School
Monterey, California 93943

Abstract: Machinery operating in non-stationary mode generates a signature which at each instant of time has a distinct frequency. A time-frequency domain representation is needed to characterize such signature. Pseudo Wigner-Ville distribution is ideally suited for portraying non-stationary signal in the time-frequency domain. The important parameters affecting the pseudo Wigner-Ville distribution are discussed and sensitivity analyses are also performed. Practical examples are also presented.

Key Words: Pseudo Wiger-Ville distribution; non-stationary phenomena; transient machinery; condition monitoring; time-frequency domain; signal processing; analytic signal; FFT; Hilbert transform; Hamming window; Gaussian window

Introduction: The physical condition or state of health of machineries which operate in transient or other non-stationary modes are difficult to predict with any degree of accuracy. It is common to practice periodic preventive maintenance on these machineries in order to avoid failures and prolong the useful operating life of the equipment.

In order to assess the physical condition of machinery without complete disassembly, a physical measurement of its vibrations is conducted using an accelerometer. Other sensors, such as temperature or pressure transducers, could also be used. There are other methods, including motor current signature analysis on electrically driven machinery and wear debris analysis which could be used. However, vibrations are used predominantly for machinery condition monitoring. The vibrations are recorded in the time domain.

There is a need for a method to represent the time dependent events which occur with machinery operating in non-stationary modes. At each instant in time as the speed of the machinery changes, the frequency content will also change. The pseudo Wigner-Ville distribution is the method which was chosen to portray these time dependent changes. This is a continuation of work initially performed and published by Rossano, Hamilton and Shin [1].

Pseudo Wigner-Ville Distribution; Analysis of Time-varying Signal:
The pseudo Wigner-Ville distribution is a three dimensional (time, frequency,

235

amplitude) representation of an input signal and is ideally suited for describing transient or other non-stationary phenomena. The Wigner Distribution (WDF) has been used in the areas of optics [2,3,4] and speech analysis [5,6]. Wahl and Bolton [7] used it to identify structure-borne noise components. Flandrin et. al. [8] recently proposed its use in the area of machinery condition monitoring and diagnostics, while Forrester [9] is investigating its use in gear fault detection.

For such a non-stationary signal analysis, spectrogram is commonly used, which is based on the assumption that it is a collection of a short duration stationary signals. A major drawback of this approach is that the frequency resolution is directly affected by the duration of short stationary time, which subsequently determines the time resolution. A method for time-frequency domain signal characterization that overcomes this drawback is the Wigner distribution which was first introduced by Wigner [10] in 1932 to study the problem of statistical equilibrium in quantum mechanics. The frequency and time resolutions of the Wigner distribution are not determined by the short duration but rather determined by the selection of desired resolution of the signal itself.

The general expression of the time-frequency distribution of a signal, $w(t,\omega)$ is given by, [11]

$$w(t,\omega) = \frac{1}{2\pi}\iiint e^{-j\theta t - j\tau\omega - j\theta u}\phi(\theta,\tau)\, s^*(u-\frac{\tau}{2})\, s(u+\frac{\tau}{2})\, du\, d\tau\, d\theta \qquad (1)$$

where $s(u)$ is the time signal, $s^*(u)$ is its complex conjugate, and $\phi(\theta, \tau)$ is an arbitrary function called the kernel. By choosing different kernels, different distributions are obtained. Wigner distribution is obtained by taking $\phi(\theta,\tau) = 1$. The range of all integrations is from $-\infty$ to ∞ unless otherwise noted.

Substituting the kernel $\phi(\theta,\tau) = 1$ to Eq. (1), the Wigner distribution is obtained,

$$w(t,\omega) = \int s^*(t-\frac{\tau}{2})\, s(t+\frac{\tau}{2})\, e^{-j\tau\omega}\, d\tau \qquad (2)$$

One of the basic frequency representations of a signal is the power density spectrum, which characterizes the signal's frequency component distribution. The power spectral density function $p(\omega)$ of a signal $s(t)$ can be related to the Fourier transform of the signal's autocorrelation function $R(\tau)$:

$$p(\omega) = \int e^{-j\omega\tau}R(\tau)\, d\tau \qquad (3)$$

with

$$R(\tau) = \int s(t)\, s(t+\tau)dt \qquad (4)$$

From this relation a time-dependent power spectral density function can be written as

236

$$w(t,\omega) = \int R_t(\tau)\, e^{-j\omega\tau}d\tau \tag{5}$$

where now $R_t(\tau)$ is a time-dependent or local autocorrelation function. Mark [12] argued for symmetry,

$$R_t(\tau) = s^*(t - \frac{\tau}{2})\, s(t + \frac{\tau}{2}) \tag{6}$$

which gives the Wigner distribution function.

Properties of Wigner Distribution Function (WDF): The properties of the WDF [13,14] are summarized and reinterpreted with this new formulation as follows: (i) the WDF is a real-valued function; (ii) the integral of the WDF with respect to frequency and time yields the instantaneous signal power and the signal's power spectral density respectively; (iii) a time or frequency shift in the signal has the same shift in the WDF; (iv) the WDF is symmetrical in time for a given signal; (v) the WDF is not always positive; (vi) the integration of the square of the WDF equals the square of the time integration of the signal's power.

Calculation with Digital Signal Processing: There are two distinct advantages for the calculation of the WDF. First, it has the form of the Fourier transform and the existing FFT algorithm can be adapted for its computation. Second, for a finite time signal, its integration is finite within the record length of the existing signal.

The discrete time Wigner distribution as developed by Claasen and Mecklenbrauker [13] is expressed by,

$$w(t,\omega) = 2 \sum_{\tau=-\infty}^{\tau=\infty} e^{-j2\omega\tau} s(t+\tau)\, s^*(t-\tau) \tag{7}$$

The discrete version of Eq. (7) for a sampled signal s(n), where n=0 to N-1, has the form,

$$w(\ell,k) = \frac{1}{N} \sum_{n=0}^{N-1} s(\ell+n)\, s^*(\ell-n)\, e^{-j\frac{4\pi}{N}nk}, \quad k=0,1,2,...N\text{-}1 \tag{8}$$

where s(m)=0 for m < 0 and m > N-1. However, in order to utilize the FFT algorithm, it must be assumed that the local autocorrelation function has a periodicity of N. This is just for operational convenience and should not apply to the interpretation of s(m). Eq. (8) can be rewritten as,

$$w[\ell, k + m(N/2)] = \frac{1}{N} \sum_{n=0}^{N-1} s(\ell+n)\, s^*(\ell-n)\, e^{-j\frac{4\pi}{N}n(k+m\frac{N}{2})} \tag{9}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} s(\ell + n)\, s^*(\ell - n)\, e^{-j\frac{4\pi}{N}nk}\, e^{-jmn2\pi}$$

$$= w(\ell, k)$$

since $e^{-jmn2\pi} = 1$ for m=integers.

Eq. (9) indicates that the WDF has a periodicity of N/2. Hence, even when the sampling of s(t) satisfies the Nyquist criteria, there are still aliasing components in the WDF. A simple approach to avoid aliasing is to use an analytic signal before computing the WDF. In 1948, J. Ville [15] proposed the use of the analytic signal in time-frequency representations of a real signal. An analytic signal is a complex signal which contains both real and imaginary components. The imaginary part is obtained by Hilbert transform. The analytic signal may be expressed by,

$$s(t) = s_r(t) + j\, H\{s_r(t)\} \tag{10}$$

where $H\{s_r(t)\}$ is a Hilbert transform and generated by the convolution of the impulse response h(t) of a 90-degree phase shift as follows:

$$H\{s_r(t)\} = s_r(t) * h(t)$$

$$h(t) = \frac{2\sin^2(\pi t/2)}{\pi t}, \qquad t \neq 0, \tag{11}$$

$$= 0, \qquad t = 0$$

where * denotes the convolution. Rewriting Eq. (11) to discrete form,

$$H\{s_r(n)\} = \sum_{m=-\infty}^{\infty} h(n-m)\, s_r(m) \tag{12}$$

The distribution resulting from an analytic signal being processed through the Wigner distribution is commonly termed as Wigner-Ville distribution.

To calculate the Wigner distribution of the sampled data, it is necessary that Eq. (8) be modified to Eq. (13), because the WDF has N/2 periodicity.

$$w(m\Delta t, k\Delta\omega) = 2\Delta t \sum_{n=0}^{2N} s[(m+n)\Delta t]\, s^*[(m-n)\Delta t]\, e^{-j2\pi nk/(2N)} \tag{13}$$

where $\Delta\omega = \pi/(2N\Delta t)$ and Δt is the sampling interval. The algorithm used in this paper is based on one written by Wahl and Bolton[7] and can be expressed as:

$$w(m\Delta t, k\Delta\omega) = \text{Re}\,[2\Delta t\,\text{FFT}(\text{corr}(i))]$$

$$\text{corr}(i) = s(m+i-1)\,s^*(m-i+1), \quad m \geq i \qquad (14)$$

$$= 0, \qquad\qquad\qquad m < i$$

where $\quad 1 \leq i \leq N+1$,

$$\text{corr}(2N-i+2) = \text{corr}^*(i), \qquad 2 \leq i \leq N$$

The frequency resolution, $\Delta\omega$, in Eq. (13) is different from that obtained by FFT of the original N point time record in two respects. The first difference is that the argument of the time signal and its conjugate contains a factor of 1/2, and secondly, the autocorrelation of the time signal is twice the length of the original record and therfore the FFT is evaluated over 2N points. The result is, that the WDF frequency resolution is one forth the resolution of an ordinary power spectrum density function.

Before processing the WDF, a modified Hamming window is applied to the time domain signal to reduce the leakage caused by the discontinuity of the finite record of data, which will be called as data tapering. This type of window is preferable since it alters the amplitude of fewer data points at the beginning and the end of the data block. A modified Hamming window, D(t) is given by:

$$D(t) = \begin{cases} 0.54 - 0.46 * \cos(10\pi t/T), & 0 \leq t \leq T/10, \\ 1.0, & T/10 \leq t \leq 9T/10, \\ 0.54 - 0.46 * \cos(10\pi(T-t)/T), & 9T/10 \leq t \leq T. \end{cases} \qquad (15)$$

Two other characteristics of the WDF should be also noted. First, the WDF of the sum of two signals is equal to the sum of the WDF of each signal plus cross term that appear when the cross-correlation of the two signal is non-zero. Second, the WDF may have negative values, which may be largely caused by interference due to the presence of these cross terms. In the case of input signals that contain multi-frequency components, the Wigner-Ville distribution of most signals are very complicated and difficult to interpret.

There are two methods to suppress the interference components of the WDF. Claasen and Mecklenbräuker[12] describe the application of a sliding window in the time domain before calculating WDF. The WDF obtained with a window function is called the Pseudo- Wigner distribution function. A second option is to smooth the WDF with a sliding averaging window in time-frequency plane. In both case the result is to deemphasize components arising from calculations and to emphasize deterministic components. Obviously, averaging a Wigner-Ville distribution will result in a Pseudo Wigner-Ville distribution.

In this research, a sliding exponential window in the time-frequency domain was chosen. That is, a Gaussian window function, G(t, w) is selected to reduce the interference and to avoid the negative values as follows:

let

$$G(t,\omega) = \frac{1}{2\pi\sigma_t\sigma_\omega} \, e^{-(\frac{t^2}{2\sigma_t^2} + \frac{\omega^2}{2\sigma_\omega^2})}, \tag{16}$$

then

$$w(t,\omega) = \frac{1}{2\pi}\iint w(t',\omega') \, G(t-t',\omega-\omega') \, dt' \, d\omega' \quad >0 \tag{17}$$

where σ_t, $\sigma_\omega > 0$ and $\sigma_t\sigma_\omega \geq 1/2$ [16]. The time and frequency resolution's Δt and $\Delta\omega$ of this Gaussian window are related by,

$$\sigma_t = j\,\Delta t, \quad \sigma_\omega = k\,\Delta\omega \tag{18}$$

in the discrete form. Then the condition for the WDF to be positive in this case is

$$j\,\Delta t\, k\,\Delta\omega \, > 1/2. \tag{19}$$

This is the time-frequency version of Heisenberg's uncertainty relation[14]. If the segmentation of time and frequency for a given signal from Eq. (2) violates this uncertainty principle, the corresponding WDF may not be positive.

To perform the convolution on the sampled WDF, the Gaussian window function was applied to the range $\pm 2\sigma_t$ and $\pm 2\sigma_\omega$. Selecting w and t to be the multiple of time and frequency steps, the sampled Gaussian window function is expressed by,

$$G(p,q) = \frac{1}{2\pi\,j\,k\,\Delta t\Delta\omega} \, e^{-\left(\frac{p^2}{2j^2} + \frac{q^2}{2k^2}\right)} \tag{20}$$

where p and q are an integer numbers in the range $\pm 2j$ and $\pm 2k$, respectively.

The convolution of the sampled WDF and the Gaussian window function can be evaluated as follows:

$$w'(\ell,m) = \frac{\Delta t\,\Delta\omega}{2\pi} \sum_{p=\ell-j}^{\ell+j} \sum_{q=m-k}^{m+k} w(p,q) \, G(p-\ell,\, q-m) \tag{21}$$

where $w'(\ell,m)$ is the smoothed WDF or Pseudo Wigner-Ville distribution.

Figure 1 shows a block diagram for computational algorithm of the Pseudo Wigner-Ville distribution. A time-varying signal sampled with the Nyquist rate is first high passed through a digital filter if the signal involves the zero frequency component, i.e., DC component, and converted into the analytic signal through a Hilbert transform. Then, the time-dependent correlation

function is computed and the result is the WDF in terms of both time and frequency domain by FFT. The final step is to compute the convolution with a Gaussian window.
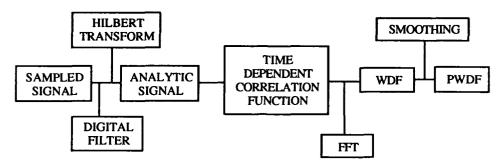
Fig. 1 Computational block diagram of Pseudo Wigner-Ville Distribution

Examples and Discussions: Machinery operating in transient mode generates a signature in which the frequency content varies at each instant of time. To characterize such signatures and to understand the vibrational behavior of such machineries, time-frequency domain representation of the signal is needed. As discussed in the previous sections, Wigner distribution is a signal transformation that is particularly suited for the time-frequency analysis of nonstationary signals. There are many advantages of using Pseudo Wigner-Ville Distribution (PWVD) for both steady and transient signals. However, there are also several disadvantages, for example, the drastic increase of peak value when the frequency content of signal changes abruptly. A computer program has been developed for PWVD and continuously updated[18]. Two different versions are available at the present time; workstation and IBM PC compatible.
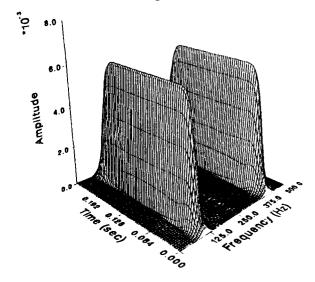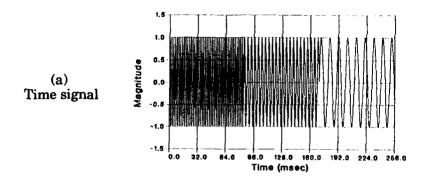
Fig. 2 Pseudo Wigner-Ville distribution of 100 and 400 Hz Pure Sine Waves
(f$_s$=1000 Hz, N=256 and Smoothing Window Size=10x10)

241

Pure Sine Wave: Figure 2 shows the PWVD of the pure sine wave with two frequency components (100 Hz, 400Hz), respectively. The modified Hamming window was applied to the time domain signal and the Gaussian smoothing window function was applied on time-frequency domain Winger-Ville distribution. The slope of the end edges are due to data tapering by using the modified Hamming window. The notation , f_s and N used in the Figures are sampling frequency and the total number of time data points.

Pure Sine Wave with Stepwise Frequency Changes: Figure 3 shows (a) the sine wave with stepwise frequency changes, 100 Hz, 250 Hz and 500 Hz and (b) its PWVD. The PWVD shows the time delay and frequency component of the signal. The wide spread of PWVD at the edge of each frequency region is noticed. This phenomenon is caused by the discontinuity of the signal in

(a)
Time signal

(b)
PWVD
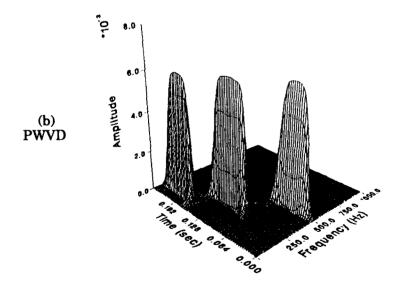
Fig. 3   Sine Wave with Stepwise Frequency Changes: 100, 250 and 500 Hz
(f_s=2000 Hz, N=512 and Smoothing Window Size=10x10)

242
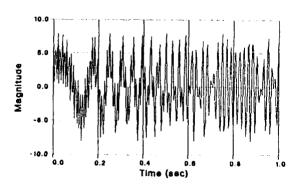
time domain and the leakage in digital signal processing. This effect may be reduced by applying the data tapering to the actual signal block. Nevertheless the PWVD represented the characteristics of the signal well. PWVD can portray the characteristics of the steady state signals involving time delay and multi-frequency components . If different size of the smoothing window are applied, the PWVD amplitude changes, but the total energy remains unchanged.

Composite Signal with Two Frequency Components at Each Time: The PWVDs of the nonstationary signals were studied and the results were shown in Figures 4 through 7. Figure 4 shows (a) the time signal composed of two sweeping frequency components at each time, one increasing and the other decreasing with the same rate, and (b) its Wigner-Ville distribution (before applying the smoothing window) and (c) its pseudo Wigner-Ville distribution (after applying the smoothing window), respectively.

The effect of cross (or interference) term is significant and appeared in the average frequency region. This is one of the disadvantages of using Wigner-Ville distribution but it is a characteristic of the distribution. When Gaussian window was applied to Wigner-Ville distribution, the effect of cross term disappeared. The main lobe of PWVD is wider and its amplitude is significantly reduced. The large peak at the intersection point of two sweeping frequency signals is mainly caused by the doubling effect of amplitudes of two signals.

A Linear Chirp Signal: Another type of a non-stationary signal sweeps up and down in frequency is called a linear chirp signal and is shown in Figure 5(a). This signal has only one frequency component at each time. The effect of cross terms appears in the Wigner-Ville distribution, as shown in Figure 4(b). The smoothing window was applied to Wigner-Ville distribution and the result is shown in Figure 5(c). As expected, the effect of cross term is significantly reduced. However, the unusual peak (called 'ghost' peak) appeared at the point where the direction of sweep changes. To understand the cause of this phenomenon, the PWVD was integrated along the frequency axis and it was found that the square root of the resultant amplitude was the amplitude of original time signal, implying that the energy content remained constant. The following function was used to generate the linear chirp signal:

(a)
Time Signal

243

(b)
WDF

(c)
PWVD
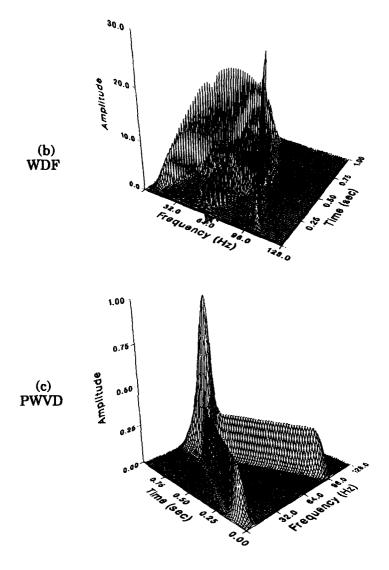
Fig. 4   Composite Signal with Two Frequency Components at Each Time;
$$s(t)=4\cos(2\pi\ 32t^2) + 4\ \cos\{2\pi(40+32(2-t)]t\}$$
(f_s=256 Hz, N=256 and Smoothing Window Size=10x10)

$$s(t) = \sin\left[2\pi\left(30 + \frac{220(i-1)}{256}\right)t\right], \qquad\qquad 1 \leq i \leq 256$$

$$s(t) = -\sin\left[2\pi\left(30 + \frac{220(512-i)}{256}\right)(0.256-t)\right], \qquad 256 \leq i \leq 512$$

where t(i-1) dt and dt=0.0005.

(22)

244

(a)
Time signal

(b)
WDF

(c)
PWVD
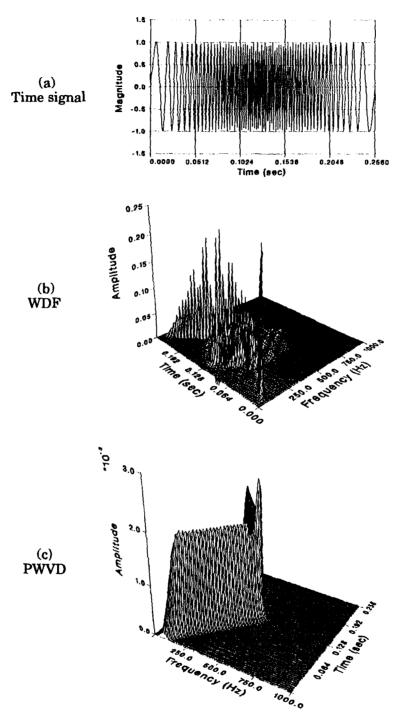
Fig. 5   Linear Chirp Signal with One Frequency Component at Each Time
(f$_s$=2000 Hz, N=512 and Smoothing Window Size=16x16)

245

A Composite Signal of Sweeping-up and Steady Frequency: The signal which sweeps up along the frequency for first 0.5 second and holds to a constant frequency for next 0.5 second was considered. This signal is typical speed profile of start-up stage of pump. Figure 6 shows (a) PWVD and (b) its contour plot. The interesting phenomenon was observed in PWVD that the sweep-up portion of signal (first half seconds) has a lower amplitude and wider main lobe compared with the steady frequency region of signal (second half seconds). When the PWVD was integrated along the frequency axis and it was found that the resultant amplitudes in these two regions are same. The following functions were used to generate the desired signal:

$$s(t) = 4\cos(2\pi 32t^2), \qquad 0 \le t \le 0.5 \text{ sec.}$$
$$s(t) = 4\cos(2\pi 64t), \qquad 0.5 \le t \le 1.0 \text{ sec.}$$

(23)

(a)
PWVD

(b)
Contour
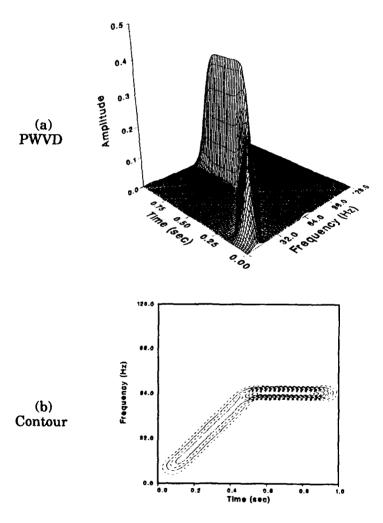
Fig. 6   PWVD of a Composite Signal of Sweeping-up and Steady Frequency
(f_s=256 Hz, N=256 and Smoothing Window Size=10x10)

246

(a) $s(t) = \cos(2\pi\ 60\ t)$

(b) $s(t) = \cos(2\pi\ 32\ t^2)$

(c) $s(t) = \cos(2\pi\ 64\ t^2)$
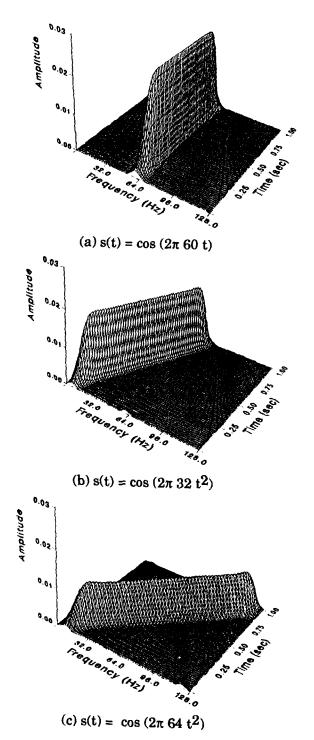
Figure 7. The Effect of Sweep Rates To Pseudo Wigner-Ville Distribution (f_s=256 Hz, N=256 and Smoothing Window Size=10x10)

247

Sweep Rate Effect: The effect of sweep rate on PWVD was investigated. The sweep rate is the frequency change per unit time. Figure 7 shows the PWVDs of the linear chirp signal with a various sweep rates:(a) has zero sweep rate and (b) has lower sweep rate than (c). It can be seen that the amplitude of PWVD decreases with increasing sweep rate but energy remains unchanged. This result appeared to be caused by Heisenberg's uncertainty relation between time and frequency. However, based on this study, it is clear that the 'ghost' peak (see Figure 5) appears due to the instantaneous zero sweep rate at the point where the direction of sweep changes. Also the peak value is affected by the size of smoothing window.

Actual Pump Start-up RPM Signal: The start-up transient speed of the pump was measured and the results were shown in Figure 9. The PWVD is shown in Figure 9(a) and the contour view is shown in Figure 9(b). The contour plot shows that the speed of the pump runs up when initially started, reaches the maximum RPM and coasts down gradually. Near the maximum speed during the run up, the sweep rate was rapidly decreased and, as a result, the peak value was rapidly increased. When the sweep rate is close to zero at the normalized time of 0.4, the amplitude attains the maximum value.

(a)
PWVD

(b)
Contour
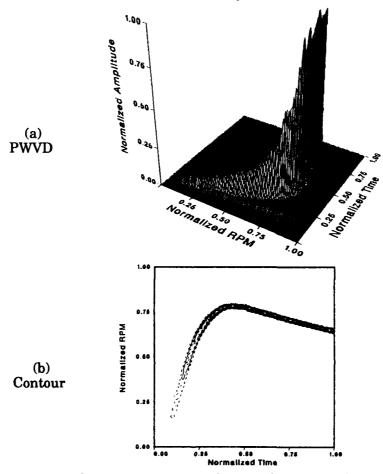
Figure 8. Pseudo Wigner-Ville Distribution of Transient Speed of the Pump

Conclusions: The pseudo Wigner-Ville distribution has been investigated and applied to analyzing non-stationary signals typical of transient machinery signatures. The results of this research will be a valuable assets for condition monitoring of transient machinery. The following conclusions can be drawn:

(1) The pseudo Wigner-Ville distribution is ideally suited for portraying non-stationary time signals.

(2) The use of modified Hamming window to time signals is effective to reduce the edge effect of discontinuity.

(3) The use of the analytic signal in calculating the Wigner distribution eliminates aliasing problem.

(4) The Gaussian window function for smoothing the Wigner-Ville distribution is very effective and the presence cross terms is significantly reduced.

(5) Both the amplitude and the main lobe of the pseudo Wigner-Ville distribution is significantly affected by the sweep rate. As the absolute sweep rate increases, the amplitude of the PWVD decreases and the main lobe becomes wider.

REFERENCES:

1. Rossano, G.W., Hamilton, J.F., and Shin, Y.S., "The Pseudo Wigner-Ville Distribution as a Method for Machinery Condition Monitoring of Transient Phenomia", Proceedings of the 2nd International Machinery Monitoring & Diagnostics Conference, Los Angeles, CA, pp. 167-173, October 22-25, 1990.

2. Bastiaans, M.J., "The Wigner Distribution Function Applied to Optical Signals and Systems." Optics Communications, vol. 25, no. 1, pp. 26-30, April 1978.

3. Bastiaans, M.J., "Wigner Distribution Function and its Application to First-Order Optics," Journal of the Optical Society of America, vol. 69, no. 12, pp. 1710-1716, Dec 1979.

4. Bartelt, H.O., Brenner, K.H. and Lohmann, A.W., "The Wigner Distribution Function and its Optical Production," Optics Communications, vol. 32, no. 1, pp. 32-38, Jan 1980.

5. Riley, M., Speech Time-Frequency Representations, Kluwer Academic Publishers, 1989.

6. Velez, E. and Absher, R., "Transient Analysis of Speech Signals using the Wigner Time-Frequency Representation," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2242-2245, May 1989.

7. Wahl, T., and Bolton, J., "The Use of the Wigner Distribution to Analyze Structural Impulse Responses," International Congress on Recent Developments in Air and Structure-Borne Sound, Mar 1990.

8. Flandrin, P., Garreau, D., and Puyal, C., "Improving Monitoring of PWR Electrical Power Plants 'In Core' Instrumentation with Time-Frequency Signal Analysis," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2246-2249, May 1989.

9. Forrester, B., "Analysis of Gear Vibration in the Time-Frequency Domain," Proceedings of the 44th Meeting of the Mechanical Failures Prevention Group, pp. 225-234, Apr 1990.

10. Wigner, E., "On the Quantum Correction for Thermodynamic Equilibrium", Physics Review, vol. 40, pp. 749-759, June 1932.

11. Cohen, Leon, "Time-Frequency Distribution - A Review," Proceeding of the IEEE, Vol.77, No.7, pp.941-981, 1989.

12. Mark, W. D., "Spectral Analysis of the Convolution and Filtering of Non-Stationary Stochastic Processing," Journal of Sound and Vib., Vol.11,pp.19-63, 1970.

13. Claasen, T. and Mecklenbrauker, W., "The Wigner Distribution - A Tool for Time-Frequency Signal Analysis," Philips Journal of Research, Vol.35, Part I: pp.217-250, Part II: pp.276-300, Part III: pp.372-389, 1980.

14. Yen, N., "Time and Frequency Representation of Acoustic Signals by means of the Wigner Distribution Function: Implementation and Interpretation," Journal of Acoustical Society of America, Vol.81, No.6, pp.1841-1850, 1987.

15. Ville, J., "Theorie et Applications de la Notion de Signal Analytique," Cables et Transmission, Vol. 2a, No.1, pp.61-74, 1948.

16. Cartwright, N. D., "A Non-Negative Wigner-Type Distribution," Physica, 83A, pp.210-212, 1976.

17. Otens, R. K., and Enochson, L., Applied Time Series Analysis, John Wiley and Sons, Chap.4, 1978.

18. Jeon, J. J. and Shin, Y.S., "Pseudo Wigner-Ville Distribution, Computer Program, and its Applications To Time-Frequency Domain Problems", NPS Technical Report (In Preparation).

TRAINING NEURAL NETWORKS
WITH
SIMULATED VIBRATION DATA TO IDENTIFY
MACHINERY FAULTS

Richard Smith
Peter Hayes & Thomas Walter

Mechanical Technology Inc.
968 Albany Shaker Road
Latham, N.Y. 12110
(518) 785-2211

Abstract: Neural networks require training with the full range of input values they will encounter in use. In machinery diagnostics, it is not practical to implement comprehensive sets of faults for training - or to wait for them to develop naturally in a machine.

This paper de ribes an initial attempt to use the human diagnostician's knowledge of a machine's symptoms/faults to a train a neural network. The approach is as follows: Based on his knowledge, the diagnostician develops synthesized samples of data which are used to train a neural network. In this case, the fault simulated represents the vibration associated with deteriorating rolling element bearings. The data consisted of simulated time based vibration signatures typically found in machines as their bearings degrade. This data was processed with a commercially available neural net software package. As a check on the predictive ability of the neural network, results obtained were compared to those offered by linear regression methods.

The work conducted concludes that simulated data can be used to train a neural network. In some cases the network outperformed standard linear regression techniques. The degree of success depends upon the completeness of and variation within the data used to train the network.

Key words: Bearings; defect detection; demodulation; diagnostics; envelope detection; machinery; monitoring; neural networks; rolling element bearings; spalls; statistics; statistical analysis; vibration.

Introduction: Vibration associated with rolling element bearings is often masked by background noise or other machine vibration. This can make interpretation of bearing vibration information extremely difficult. Traditional diagnostic methods have concentrated on processing vibration signals by enhancing specific defect related features. Filtering and/or other data screens

251

are commonly used, along with signal level detectors. Times series averaging, spectral processing and peak height detection are often successful [1]. Bearing analyzers use measures of spike energy, shock pulse, crest factors, and envelope demodulation to enhance frequencies of interest [2].

Advances in computer processing speed and power are now providing opportunities for additional diagnostic assessment of machinery vibration signatures. Expert Systems, Artificial Intelligence, Fuzzy Logic, and Neural Networks are potentially powerful techniques. Before Neural Networks can be practically applied, however, a way must be found to train them without implementing faults in machines, or waiting for them to occur naturally.

This paper relates the results of an initial effort to train a network by directly using the diagnostician's knowledge of how degrading components manifest themselves in observable symptoms. This was done by developing synthetic "wave forms" and using them to train a network. In addition, a statistical analysis software package [3] was used as an independent "check" of the neural network's predictive ability.

Summary of Work: Our goal was to train a neural network to identify pulses in time domain data that are representative of the rapid amplitude changes observed in deteriorating rolling element bearings. Since neural networks must be trained to identify anticipated machine based signals before they can identify incoming defect information, a range of pulse data was generated for presentation to the network. A software neural network [4] was used on a 386-20 MHZ computer with math co-processor. Various combinations of layers and nodes were used in training, but the most common configuration consisted of 10 input nodes, 5 hidden layers and 3 output nodes. A back propagation builder was used to assemble the initial network.

Digitized simulation vibration data was used. Figure 1 illustrates a typical sequence of the simulated data used to train and test the network. The simulated data was the equivalent of vibration signatures that are commonly emitted by rolling element bearings as they develop internal spalls and wear debris. Amplitudes below a selected level were designated as noise. The synthesized data was generated with computerized mathematical algorithms, and represented actual machinery time-based vibration. The advantage of using generated signatures for training was the control over the distribution and amplitude of the peaks, and the degree of background noise that could be presented to the network. The simulated data provided an opportunity for evaluating the neural network response over a full range of anticipated signatures. The pattern of the signatures ranged from fully cyclic in nature to fully random. The simulated vibration amplitudes varied nominally from 0 to .8. The training data was intended to represent the range of vibration data that might be encountered in a machine.

If the number of available modeling parameters is too high relative to the amount of incoming data (say a parameter for every pair of data points), the network may completely explain the data. It is the authors' opinion that this is why some papers in current literature offer such good results for some neural networks. The training data in the present study was increased in volume until neither the neural network nor the statistical technique used as a comparison was over-specified. Data consisted of 7000+ rows with 10 or more data points per row.

Within the training data, there were usually several data points in the noise region for every point above the noise. Figure 1 presented the relative distribution of noise and pulses used in a set of data. Figure 2 shows the typical association between adjacent points in the data. The figure shows the relative occurrence of adjacent peaks.

Figure 3 shows the values provided by the neural network from a set of test data. The outputs given by the network were associated with the three possible decisions the network had to make concerning the input data (i.e., the output nodes); "Is a Peak", "Maybe a Peak" and "Not a Peak". The output data fell into two separate planes of output values. A standard multi-variant regression fit of a similar but larger data set is shown in Figure 4. The two analysis methods properly predicted a peak in about 75% of the test cases.

Neural Network Vs Linear Regression Techniques: The predictive ability of a trained neural network is often judged by counting the number of correct decisions the network makes when presented with known inputs as test cases. The network provides a set of values used to generate the "hits" or "misses" in a "scoring matrix" format.

Scoring Matrix elements correspond to the plot regions in the following numbered pattern:

1	2	3
4	5	6
7	8	9

The most accurate predictions will have the highest number of points in the 1-5-9 diagonal and very few entries in the remaining elements.

It is interesting to note how the statistical approach compared to the neural network. The scoring matrix was used as the evaluation method. A sample statistical plot showing the predicted values against the actual ones desired was displayed previously in Figure 4. If there was 100% accuracy in the
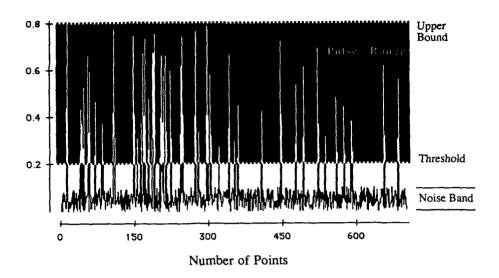
Figure 1 Simulated Time Based Signal Sample Used in Study
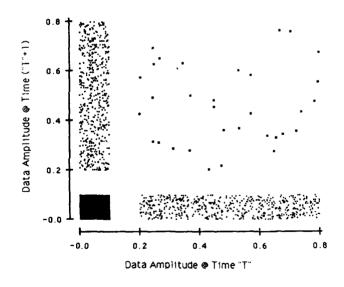
Figure 2 2-D Distribution Density Plot of One Training Data Set

254
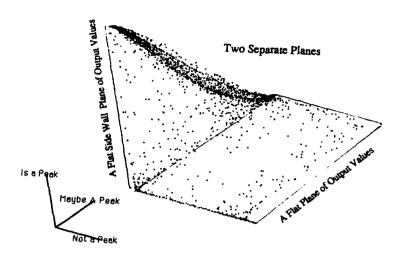
Figure 3 Neural Scalar Outputs Plotted Against Actual Values from One Data Set.
Showing Non-linear Clustering Along One Plane Diagonal.
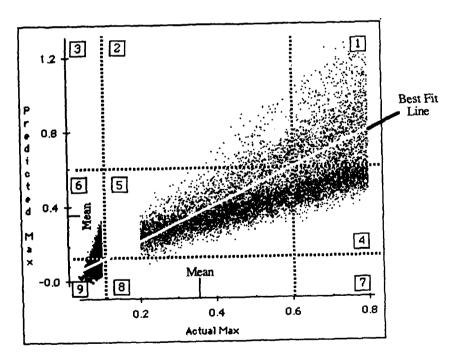
Figure 4 Multi-variable Statistical Fit from One Sample Data Set

model, all data would fall on a single straight line. Obviously this did not happen and the question is - How good was the matrix for the statistical approach? An approximate answer is represented by the nine dotted regions in Figure 4. Within the nine regions, the figure graphically approximates the nine matrix elements of the scoring matrix given in the neural network output analyses. The "Scoring Matrix" in this case is the number of data points within each of the 9 bounded regions shown in Figure 4.

Results: The study established that the neural network could settle on a solution which was not optimum, even when presented with sufficient data. A comparison of a non-optimum result is displayed in Figure 5. The figure shows a set of multi-variable fit predicted values plotted against the actual values, which in this case were used for input. Figure 6 shows the same data correlated with a neural network (R~.75 vs R~.65).

Figures 7 & 8 show a situation where the opposite was true. Figure 7 gives the results of the statistical fit. Figure 8 shows the results of using the neural network to fit the same data. The goodness of fit was better than the statistical approach (R~.98 vs R~.995). With the neural network, changing the number of hidden layers and nodes used for training provided different fit results from the same data sets.

Statistical data fitting provides an equation of estimation. The network also provided an equation for the algorithm established during the training period. This feature is very useful, since the algorithm can be used in other software as a stand-alone subroutine. These routines can then be implemented as diagnostic tools for processing test data.

The study also confirmed that even well-trained networks have difficulty predicting accurately when test data falls beyond the limits of the network's training. In practice, for example, changes in the average noise level from a sensor would degrade the network's predictive ability. If various noise levels are likely to be encountered in practice, then the network must be trained for that eventuality. This observation emphasizes the importance of using a full range of data for training.

An additional observation: One method of adding robustness to the network's operation would be to augment its inputs with statistical features that are important to the pattern being distinguished. Such features might include the average values of a block of amplitudes, the standard deviation of these amplitudes and/or the fact that an amplitude exceeds the accepted noise threshold. This approach was tried and found to significantly improve the accuracy of the trained network. Data which had a correlation coefficient of 75% between predicted and actual values could be improved to well over 95% when one or more of the data features cited were used in the analytical predictive equation.
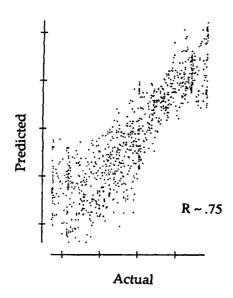
Figure 5 A Statistical Fit - Predicted Versus Actual
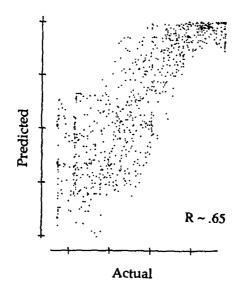
Figure 6 A Non-Optimum Neural Network Fit to the Same Data as Shown Above
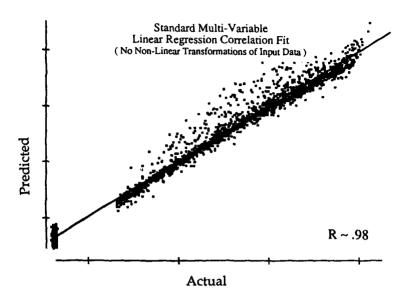
Figure 7 A Statistical Fit - Predicted Versus Actual
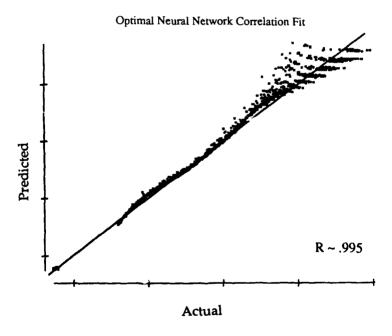
Figure 8 A Near-Optimum Neural Fit to the Same Data as Shown Above

Conclusions and Recommendations: Simulated vibration data can be used to train a neural network. The degree of success depends on 1) the trainer's ability to adequately simulate all likely scenarios, and 2) the statistical variation within the data set used to train the network.

When properly implemented and trained, the neural network used can exceed the "Bad/Good" pattern recognition accuracy of a Multi-Linear Regression equation generated from the same data set.

Additional work should be accomplished, especially using trained neural networks on "live" data, and simulation/verification of other machinery faults.

References:

1. Smith, R. L., "Rolling Element Bearing Diagnostics with Lasers, Microphones and Accelerometers", 46th MFPG Meeting April 1992.

2. Frarey, J. L. and Smith, R. L., Statistical Techniques for Detecting Bearing Defects", NASA Conference on Advanced Earth to Orbit Propulsion Technology, Marshall Space Flight Center, Huntsville, AL, May 13-15, 1986.

3. Data Desk® Program. Data Description Inc., P.O. Box 4555, Ithaca, NY, 14852.

4. NeuralWorks (NWorks) Professional II, by NeuralWare, Inc., Building IV, Suite 227, Penn Center West, Pittsburgh, PA 15276 was used for the neural network processing.

FAILURE MECHANISMS
AND
LIFE EXTENSION

Cochairmen: Lawrence J. Mertaugh
               Naval Air Warfare Center

               Robert A. Bayles
               Naval Research Laboratory

WEAR RELIABILITY AND PREVENTIVE REPLACEMENT POLICY FOR MECHANICAL COMPONENTS

Dimitri B. Kececioglu and Feng-Bin Sun
Department of Aerospace and Mechanical Engineering
The University of Arizona
Tucson, AZ 85721

Abstract: This paper explores the stochastic behavior of the wear process from the cumulative damage point of view. The general wear progression envelope is presented and the typically observed three wear periods are defined and discussed. Curvilinear-linear-curvilinear wear equations are fitted to data on both the lower and the upper boundaries of the wear data envelopes using the least-squares regression method. Parameters are estimated for the wear-life distribution families using the "3σ" theorem for the normal distribution and the matching percentiles method for the Weibull distribution. Wear reliability prediction procedures are developed for different cases using the normal and the Weibull distributions. The preventive replacement policy models are developed for the specified in-service reliability and for the minimum cost. Numerical examples are given and discussed. The methodologies presented in this paper can be applied to other failure modes exhibiting cumulative damage behaviors, such as metal fatigue, fatigue crack growth, corrosion, erosion, creep, deteriorating material properties in plastics with time, and so on.

Key Words: Cumulative damage; linear-squares regression; minimum cost; preventive replacement; reliability; stochastic process; "3σ" theorem; wear; wear-life distribution families; wear progression envelope.

Introduction: Wear is a very predominant failure mode for dynamically functioning mechanical components, such as gears, splines, seals, bearings, couplings, etc [1; 2]. Wear failures may be caused by the lack of proper lubrication, misalignment, high operating speeds, high operating temperatures, improper materials, etc. While wear failure modes are generally not catastrophic, they significantly add to the cost of maintaining the equipment and the operating system. In case of lack of regular preventive inspection, maintenance and replacement, machines may be damaged or destroyed, and even human lives may be lost. The objective of this paper is to develop generalized stochastic math models to represent the wear behavior; provide the theory to quantify the actual wear distribution for a specified operating time and quantify the actual distribution of the components' lives for a specified amount of wear; develop methodologies for predicting the wear reliability for any desired operating time or the allowable wear, or both; and provide the methodologies for preventive replacement scheduling.

Stochastic Behavior of Wear Process: Wear process is a cumulative damage process. During cyclic operation, a mechanical component operating in a certain environment experiences irreversible accumulation of damage from wear. These irreversible damages accumulate until the component can no longer perform satisfactorily. The component is then said

263
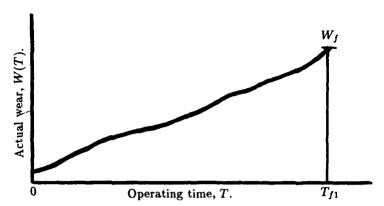
Fig. 1– The sample function (sf) of wear $W(T)$ to failure at W_f.

to have failed. The time at which the component ceases to perform satisfactorily is called the time to failure or the lifetime of the component. The process by which the irreversible damage accumulates is called a *cumulative damage (CD) process.*

Let us consider the wear of a tire on an automobile as an example. Assume the tire is initially new. The depths of the various grooves are measured at various places around and across the tire [3]. As the tire wears, there will be a gradual loss of depth of the groove; the loss in depth is an observable for this CD process. We record at various times the loss in depth at various points on the tires that are being monitored. A tire is withdrawn (or has failed) from normal service when this observable wear reaches a prescribed value at one or more of the points being monitored. Let $W(T)$ denote the loss in depth (wear) as a function of the cumulative operating time that is being used to determine when the tire is to be withdrawn from service. The value of $W(T)$ could be some average of the loss in depth at the points monitored, or it could be the largest of the losses in depth at the points being monitored, and so on. Let W_f denote the value of $W(T)$ at which the tire is withdrawn. Figure 1 shows the evolution of $W(T)$ for one tire as a function of time T. T_{f1} denotes the time at which $W(T_{f1}) = W_f$. As more tires are run, we obtain more $W(T)$ versus T curves. Figure 2 shows the $W(T)$ versus T curves for five (5) tires.

The $W(T)$ versus T curves are called *sample functions (sf's)* of this wear process. Each tire has its own sf; consequently, we have as many $sf's$ as there are tires tested. These $sf's$ are monotonically nondecreasing. The probability that two $sf's$ will coincide is negligible due to the inherent variability in service conditions and in the tire manufacturing process. Therefore, the damage (wear) levels for n different tires at a given time point T_0; i.e., $W_1(T_0), W_2(T_0), ..., W_n(T_0)$, the times to a specified damage (wear) level W_0; i.e., $T_1(W_0), T_2(W_0), ..., T_n(W_0)$, and the times to failure $T_{f1}, T_{f2}, ..., T_{fn}$ will, in general, be different and distributed.

The initial wear levels can be different due to variable manufacturing quality control of new items, variable deterioration during storage until the item is put into service, and so on.

The wear level at which failure, or retirement, occurs can arise in a number of ways and therefore may have variability. For example, a tire may be cut or punctured. A cutting tool may be considered worn out when it cuts poorly, where "poorly" determines a subinterval of values over the wear range of the tool, etc.

A typical wear process for mechanical components is shown in Fig. 3. There is a short *break-*

264
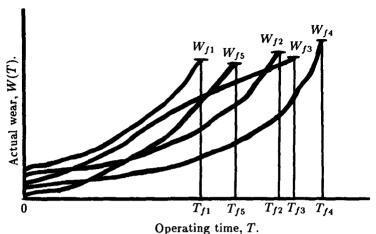
Fig. 2– Five $sf's$ of wear $W(T)$ to failure at W_f.
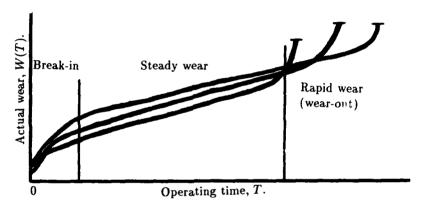
Fig. 3– A typical wear process for mechanical components.

in (early wear) period during which wear accumulates rapidly. After this initial period. wear accumulates rather steadily which is reflected in the more or less constant slope of $sf's$: this may be regarded as a *steady wear accumulation period*. Finally, in the third period, there is a rapid wear accumulation to failure, which may be regarded as the *wear-out period*. This kind of sf behavior occurs particularly in physical wear in bearings, piston rings, locks, and so on.

The stochastic behavior of the wear process can be summarized as follows:

1. The initial wear level is random.

2. The wear level at a specified time of operation is random.

3. The time to a specified wear level is random.

4. The wear level at failure is random.

5. The time to a failure wear level is random.

265
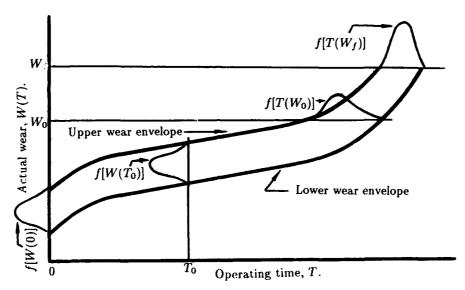
Fig. 4— Wear-life distribution families demonstrating the stochastic behavior of the wear process.

Therefore the wear process can be described by two distribution families; i.e., the distribution family of times to any specified wear level and distribution family of wear levels at any specified time of operation. These two distribution families form two envelopes; i.e., the lower wear/life limit and the upper wear/life limit envelopes. This is shown in Fig. 4. In practical situations for most mechanical components it is often much easier to obtain the plots for two (lower and upper) envelopes than to obtain the detailed plots for the two distribution families directly; i.e., distributions of wear for a specified operating time, or of lifetimes for a specified wear. However, with these two lower and upper limit envelopes, we can find the two distribution families, correspondingly, which will be explored next.

Fitting Equations to the Envelop Data: Since, Fig. 3 is the general picture of the wear process, other $sf's$ may or may not have all of the three wear periods, then they can be considered as special cases. Therefore, it is sufficient to fit equations to the envelope data of the general wear behavior. A reasonable equation for an envelope of Fig. 3 is a combination of two power functions corresponding to the first (break-in) and the third (wear-out) periods and one linear function corresponding to the second wear period (steady wear); i.e.,

$$W(T) = \begin{cases} a_0 T^{b_0} + W_0, & \text{for } 0 \leq T < T_1, \\ b_1(T - T_1) + W_1, & \text{for } T_1 \leq T < T_2, \\ a_2(T - T_2)^{b_2} + W_2, & \text{for } T \geq T_2, \end{cases} \tag{1}$$

where $a_0, b_0, W_0, b_1, W_1, a_2, b_2$, and W_2 are unknown constants while T_1 is the time point at which the first wear period ends and the second period begins, and T_2 is the time point at which the second period ends and the third period begins, as shown in Fig. 5. Note that W_0 is the initial wear; W_1 and W_2 are wears at T_1 and T_2, respectively.

Given n observations $(t_1, w_1), (t_2, w_2), ..., (t_n, w_n)$ where $t_1 < t_2 < ... < t_n$, the unknown constants can be determined by the least-squares regression technique as follows:

266
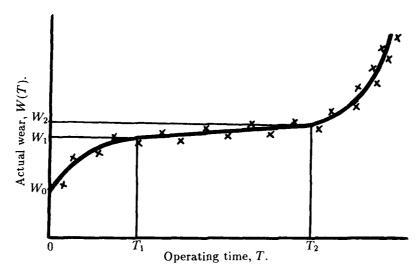
Fig. 5— One envelope of the wear process.

Step 1: By visual inspection, make an initial guess of time points T_1 and T_2, $\hat{T}_1$ and $\hat{T}_2$, which separate the three wear periods.

Step 2: Divide the n observations into three (3) groups with sample sizes of n_0, n_1 and n_2, respectively: $(t_1, w_1), (t_2, w_2), ..., (t_{n_0}, w_{n_0})$; $(t_{n_0+1}, w_{n_0+1}), (t_{n_0+2}, w_{n_0+2}), ...,$ $(t_{n_0+n_1}, w_{n_0+n_1})$; $(t_{n_0+n_1+1}, w_{n_0+n_1+1}), (t_{n_0+n_1+2}, w_{n_0+n_1+2}), ..., (t_n, w_n)$.

Step 3: The least-squares estimates for a_0, b_0 and W_0 are

$$\begin{cases} \hat{W}_0 = \frac{W_1' W_2' - (W_3')^2}{W_1' + W_2' - 2W_3'}, \\ \hat{b}_0 = \frac{L_{xy0}}{L_{xx0}}, \\ \hat{a}_0 = e^{\bar{y}_0 - b_0 \bar{x}_0}, \end{cases} \tag{2}$$

with the correlation coefficient of

$$\rho_0 = \frac{L_{xy0}}{\sqrt{L_{xx0} L_{yy0}}}, \tag{3}$$

where W_1', W_2' and W_3' are wear levels at three arbitrary time points t_1', t_2' and $t_3' = \sqrt{t_1' t_2'}$ in the first wear period (note that W_3' can be calculated by linear interpolation if t_3' is not coinciding with any one of the given observed time points), and

$$\bar{x}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} \log_e t_i, \tag{4}$$

$$\bar{y}_0 = \frac{1}{n_0} \sum_{i=1}^{n_0} \log_e (W_i - \hat{W}_0), \tag{5}$$

$$L_{xx0} = \sum_{i=1}^{n_0} (\log_e t_i - x_0)^2, \tag{6}$$

$$L_{yy0} = \sum_{i=1}^{n_0} [\log_e (W_i - \hat{W}_0) - y_0]^2, \tag{7}$$

267

and

$$L_{xy0} = \sum_{i=1}^{n_0}(\log_e t_i - \bar{x}_0)[\log_e(W_i - \hat{W}_0) - \bar{y}_0]. \tag{8}$$

Step 4: The least-squares estimates for b_1 and W_1 are

$$\begin{cases} \hat{b}_1 &= \frac{L_{xy1}}{L_{xx1}}, \\ \hat{W}_1 &= \bar{y}_1 - \hat{b}_1\bar{x}_1, \end{cases} \tag{9}$$

with the correlation coefficient of

$$\rho_1 = \frac{L_{xy1}}{\sqrt{L_{xx1}L_{yy1}}}, \tag{10}$$

where

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=n_0+1}^{n_0+n_1}(t_i - \hat{T}_1) = \frac{1}{n_1}\sum_{i=n_0+1}^{n_0+n_1}t_i - \hat{T}_1 = \bar{t}_1 - \hat{T}_1, \tag{11}$$

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=n_0+1}^{n_0+n_1}W_i = \overline{W}_1, \tag{12}$$

$$L_{xx1} = \sum_{i=n_0+1}^{n_0+n_1}(t_i - \bar{t}_1)^2, \tag{13}$$

$$L_{yy1} = \sum_{i=n_0+1}^{n_0+n_1}(W_i - \overline{W}_1)^2, \tag{14}$$

and

$$L_{xy1} = \sum_{i=n_0+1}^{n_0+n_1}(t_i - \bar{t}_1)(W_i - \overline{W}_1). \tag{15}$$

Step 5: The least-squares estimates for a_2, b_2 and W_2 are

$$\begin{cases} \hat{W}_2 &= \frac{W_1''W_2'' - (W_3'')^2}{W_1'' + W_2'' - 2W_3''}, \\ \hat{b}_2 &= \frac{L_{xy2}}{L_{xx2}}, \\ \hat{a}_0 &= e^{\bar{y}_2 - \hat{b}_2\bar{x}_2}, \end{cases} \tag{16}$$

with the correlation coefficient of

$$\rho_2 = \frac{L_{xy2}}{\sqrt{L_{xx2}L_{yy2}}}, \tag{17}$$

where W_1'', W_2'' and W_3'' are wear levels at three arbitrary time points t_1'', t_2'' and $t_3'' = \sqrt{t_1''t_2''}$ in the third wear period (note that W_3'' can be calculated by linear interpolation if t_3'' is not coinciding with any one of the given observed time points) and

$$\bar{x}_2 = \frac{1}{n_2}\sum_{i=n_0+n_1+1}^{n}\log_e(t_i - \hat{T}_2), \tag{18}$$

$$\bar{y}_2 = \frac{1}{n_2}\sum_{i=n_0+n_1+1}^{n}\log_e(W_i - \hat{W}_2), \tag{19}$$

$$L_{xx2} = \sum_{i=n_0+n_1+1}^{n}\left[\log_e(t_i - \hat{T}_2) - \bar{x}_2\right]^2, \tag{20}$$

$$L_{yy2} = \sum_{i=n_0+n_1+1}^{n}\left[\log_e(W_i - \hat{W}_2) - \bar{y}_2\right]^2. \tag{21}$$

268

and

$$L_{xy2} = \sum_{i=n_0+n_1+1}^{n} \left[\log_e(t_i - \hat{T}_2) - \bar{x}_2\right]\left[\log_e(W_i - \hat{W}_2) - \bar{y}_2\right]. \tag{22}$$

Step 6: Check the values of $\hat{T}_1$ and $\hat{T}_2$ to see if

$$\hat{a}_0\hat{T}_1^{\hat{b}_0} + \hat{W}_0 = \hat{W}_1, \tag{23}$$

and

$$\hat{b}_1(\hat{T}_2 - \hat{T}_1) + \hat{W}_1 = \hat{W}_2, \tag{24}$$

or equivalently check if the values of $\hat{T}_1$ and $\hat{T}_2$ meet the following relationships:

$$\hat{T}_1 = \left(\frac{\hat{W}_1 - \hat{W}_0}{\hat{a}_0}\right)^{1/\hat{b}_0}, \tag{25}$$

and

$$\hat{T}_2 = \frac{\hat{W}_2 - \hat{W}_1}{\hat{b}_1} + \hat{T}_1. \tag{26}$$

If they do, then the determined envelope equations are acceptable. If not, then in Step 1 use the $\hat{T}_1$ and $\hat{T}_2$ values calculated by Eqs. (25) and (26) and repeat Steps 2 through 6 until Eqs. (23) and (24) are satisfied. Equations for another envelope can be similarly determined.

Determination of the Wear-life Distribution Families: Once each wear process envelope or Eq. (1), is obtained, the distribution of the wear level at any specified (cumulative) operating time, and the distribution of the time to the specified wear level, can be quantified. The normal and Weibull distribution families are used to represent the distributions of the wear at specific lives; and of the life at specific wear levels. The applications of these two distributions are discussed next.

Fitting the Normal Distribution – The "3σ" Theorem: Assume both the wear level distribution at a specified operating time T_0, $f[W(T_0)]$, and the distribution of time to a specified wear level W_0, $f[T(W_0)]$, are normal; i.e.,

$$f[W(T_0)] = \frac{1}{\sqrt{2\pi}\sigma_{W(T_0)}}e^{-\frac{1}{2}\left[\frac{W(T_0)-\mu_{W(T_0)}}{\sigma_{W(T_0)}}\right]^2}, \tag{27}$$

and

$$f[T(W_0)] = \frac{1}{\sqrt{2\pi}\sigma_{T(W_0)}}e^{-\frac{1}{2}\left[\frac{T(W_0)-\mu_{T(W_0)}}{\sigma_{T(W_0)}}\right]^2}, \tag{28}$$

where

$\mu_{W(T_0)}, \sigma_{W(T_0)}$ = mean and standard deviation of the wear at time T_0, respectively,

and

$\mu_{T(W_0)}, \sigma_{T(W_0)}$ = mean and standard deviation of the time to the wear level W_0, respectively.

269

Then, it is reasonable to think of the lower envelope as the "$\mu - 3\sigma$" limit (or 0.135% percentage point) and the upper envelope as the "$\mu + 3\sigma$" limit (or 99.865% percentage point) of the normal wear and life distribution families, since 99.73% of the whole distribution is covered within the range of $(\mu - 3\sigma, \mu + 3\sigma)$. Therefore, $\mu_{W(T_0)}, \sigma_{W(T_0)}, \mu_{T(W_0)}$ and $\sigma_{T(W_0)}$ can be calculated by

$$\begin{cases} \mu_{W(T_0)} &= \frac{W_u(T_0) + W_l(T_0)}{2}, \\ \sigma_{W(T_0)} &= \frac{W_u(T_0) - W_l(T_0)}{6}, \end{cases} \tag{29}$$

and

$$\begin{cases} \mu_{T(W_0)} &= \frac{T_u(W_0) + T_l(W_0)}{2}, \\ \sigma_{T(W_0)} &= \frac{T_u(W_0) - T_l(W_0)}{6}, \end{cases} \tag{30}$$

where

$\qquad W_l(T_0), W_u(T_0) \quad = \quad$ lower and upper limits of wear at time T_0,

and

$\qquad T_l(W_0), T_u(W_0) \quad = \quad$ lower and upper limits of time to wear level W_0.

The values of $W_u(T_0)$ and $W_l(T_0)$ can be calculated directly by substituting $T = T_0$ into Eq. (1). The values of $T_l(W_0)$ and $T_u(W_0)$ can be obtained by substituting $W(T) = W_0$ into Eq. (1) and solving for T.

Fitting the Weibull Distribution – The Matching Percentiles Method: Assume both the wear level distribution at a specified operating time T_0, $f[W(T_0)]$, and the distribution of the time to a specified wear level W_0, $f[T(W_0)]$, are Weibull; i.e.,

$$f[W(T_0)] = \frac{\beta_{W(T_0)}}{\eta_{W(T_0)}} \left[\frac{W(T_0)}{\eta_{W(T_0)}} \right]^{\beta_{W(T_0)} - 1} e^{-\left[\frac{W(T_0)}{\eta_{W(T_0)}} \right]^{\beta_{W(T_0)}}}, \tag{31}$$

and

$$f[T(W_0)] = \frac{\beta_{T(W_0)}}{\eta_{T(W_0)}} \left[\frac{T(W_0)}{\eta_{T(W_0)}} \right]^{\beta_{T(W_0)} - 1} e^{-\left[\frac{T(W_0)}{\eta_{T(W_0)}} \right]^{\beta_{T(W_0)}}}, \tag{32}$$

where

$\qquad \beta_{W(T_0)}, \eta_{W(T_0)} \quad = \quad$ shape and scale parameters of Weibull wear distribution

$\qquad\qquad\qquad\qquad$ at time T_0, respectively,

and

$\qquad \beta_{T(W_0)}, \eta_{T(W_0)} \quad = \quad$ shape and scale parameters of Weibull distribution of time

$\qquad\qquad\qquad\qquad$ to a specified wear W_0, respectively.

Similar to the "3σ" theorem in the normal distribution case, we can think of the lower and upper limit envelopes as the 0.135% and 99.865% percentage points of the Weibull wear and life distribution families. These two percentage points may be changed to the applicable values from actual data if a greater coverage range of more than 99.73% is decided upon. Therefore, $\beta_{W(T_0)}, \eta_{W(T_0)}, \beta_{T(W_0)}$ and $\eta_{T(W_0)}$ can be obtained as follows:

$$\begin{cases} P[W \leq W_l(T_0)] &= 1 - e^{-\left[\frac{W_l(T_0)}{\eta_{W(T_0)}} \right]^{\beta_{W(T_0)}}} = 0.00135, \\ P[W \leq W_u(T_0)] &= 1 - e^{-\left[\frac{W_u(T_0)}{\eta_{W(T_0)}} \right]^{\beta_{W(T_0)}}} = 0.99865. \end{cases} \tag{33}$$

270

Solving for $\beta_{W(T_0)}$ and $\eta_{W(T_0)}$ yields

$$\begin{cases} \hat{\beta}_{W(T_0)} & = & \dfrac{8.4952}{\log_e\left[\frac{W_u(T_0)}{W_l(T_0)}\right]}, \\[3mm] \hat{\eta}_{W(T_0)} & = & W_l(T_0)\, 0.00135^{-\frac{1}{\hat{\beta}_{W(T_0)}}}. \end{cases}$$ (34)

Similarly

$$\begin{cases} P[T \leq T_l(W_0)] & = & 1 - e^{-\left[\frac{T_l(W_0)}{\eta_{T(W_0)}}\right]^{\beta_{T(W_0)}}} = 0.00135, \\[3mm] P[T \leq T_u(W_0)] & = & 1 - e^{-\left[\frac{T_u(W_0)}{\eta_{T(W_0)}}\right]^{\beta_{T(W_0)}}} = 0.99865. \end{cases}$$ (35)

Solving for $\beta_{T(W_0)}$ and $\eta_{T(W_0)}$ yields

$$\begin{cases} \hat{\beta}_{T(W_0)} & = & \dfrac{8.4952}{\log_e\left[\frac{T_u(W_0)}{T_l(W_0)}\right]}, \\[3mm] \hat{\eta}_{T(W_0)} & = & T_l(W_0)\, 0.00135^{-\frac{1}{\hat{\beta}_{T(W_0)}}}. \end{cases}$$ (36)

The meaning of $W_l(T_0), W_u(T_0), T_l(W_0)$ and $T_u(W_0)$ are the same as in normal case.

Wear Reliability Prediction: Case 1: Given the specified allowable wear, W_c, and the cumulative operating time (mission time), T_0, the wear reliability is

$$R(T_0) = P[W(T_0) \leq W_c].$$ (37)

If the normal distribution is assumed for $W(T_0)$, then

$$R(T_0) = \Phi\left[\frac{W_c - \mu_{W(T_0)}}{\sigma_{W(T_0)}}\right],$$ (38)

where

Φ = cumulative distribution function (CDF) of the standardized normal distribution $N(0,1)$.

If the Weibull distribution is assumed for $W(T_0)$, then

$$R(T_0) = 1 - e^{-\left[\frac{W_c}{\eta_{W(T_0)}}\right]^{\beta_{W(T_0)}}}$$ (39)

Note that the above calculated wear reliability can be obtained equivalently from the probability that the time the allowable wear is equal to or longer than the desired operating time, then

$$R(T_0) = P[T(W_c) \geq T_0].$$ (40)

If normal distribution is assumed for $T(W_c)$, then

$$R(T_0) = \Phi\left[\frac{\mu_{T(W_c)} - T_0}{\sigma_{T(W_c)}}\right].$$ (41)

If Weibull distribution is assumed for $T(W_c)$, then

$$R(T_0) = e^{-\left[\frac{T_0}{\eta_{T(W_c)}}\right]^{\beta_{T(W_c)}}}$$ (42)

Though the results obtained by the above two approaches should be quite close to each other, the first approach is recommended because the distribution of wear at the prespecified operating time can be determined more precisely than the distribution of the times to a prespecified amount of wear.

Case 2: Given the specified allowable wear W_c and the normally distributed duty cycle time $t \sim N(\mu_t, \sigma_t)$, then the wear reliability is

$$R(t) = P[T(W_c) - t \geq 0]. \tag{43}$$

If normal distribution is assumed for $T(W_c)$, then

$$R(t) = \Phi \left[\frac{\mu_{T(W_c)} - \mu_t}{\sqrt{\sigma^2_{T(W_c)} + \sigma^2_t}} \right]. \tag{44}$$

If Weibull distribution is assumed for $T(W_c)$, we can not get the explicit solution for $R(t)$. But numerical solution can be obtained using Monte-Carlo simulations or numerical integration [6; 7].

Case 3: Given the normally distributed allowable wear $W_c \sim N(\mu_{W_c}, \sigma_{W_c})$ and the duty cycle time (mission time) T_0, then the wear reliability is

$$R(T_0) = P[W_c - W(T_0) \geq 0]. \tag{45}$$

If normal distribution is assumed for $W(T_0)$, then

$$R(T_0) = \Phi \left[\frac{\mu_{W_c} - \mu_{W(T_0)}}{\sqrt{\sigma^2_{W_c} + \sigma^2_{W(T_0)}}} \right]. \tag{46}$$

If Weibull distribution is assumed for $W(T_0)$, then we can not get the explicit solution for $R(T_0)$. But numerical solution can be obtained using Monte-Carlo simulations or numerical integration [6; 7].

The **conditional reliability** for an additional mission time of ΔT given that the components have already satisfactorily operated for T_0 hours is

$$R(T_0, \Delta T) = \frac{R(T_0 + \Delta T)}{R(T_0)}. \tag{47}$$

If normal distribution is assumed for $W(T_0 + \Delta T)$, then

$$R(T_0 + \Delta T) = \Phi \left[\frac{\mu_{W_c} - \mu_{W(T_0 + \Delta T)}}{\sqrt{\sigma^2_{W_c} + \sigma^2_{W(T_0 + \Delta T)}}} \right]. \tag{48}$$

Therefore

$$R(T_0, \Delta T) = \frac{\Phi \left[\frac{\mu_{W_c} - \mu_{W(T_0 + \Delta T)}}{\sqrt{\sigma^2_{W_c} + \sigma^2_{W(T_0 + \Delta T)}}} \right]}{\Phi \left[\frac{\mu_{W_c} - \mu_{W(T_0)}}{\sqrt{\sigma^2_{W_c} + \sigma^2_{W(T_0)}}} \right]}. \tag{49}$$

Preventive Replacement Scheduling for the Specified In-service Reliability:

Case 1–Fixed Allowable Wear: Given the specified allowable wear, W_c, the preventive replacement time T_p for a desired in-service component reliability $R(T_p) = R_0$ can be determined as follows: Since

$$R(T_p) = P[T(W_c) \geq T_p] = R_0, \tag{50}$$

if the normal distribution is used for $T(W_c)$, then

$$1 - \Phi\left[\frac{T_p - \mu_{T(W_c)}}{\sigma_{T(W_c)}}\right] = R_0,$$

or

$$T_p = \mu_{T(W_c)} + \sigma_{T(W_c)}\Phi^{-1}(1 - R_0), \tag{51}$$

where

Φ^{-1} = inverse function of CDF of the standardized normal distribution, whose value can be found from $\Phi(z)$ tables.

If the Weibull distribution is used for $T(W_c)$, then

$$e^{-\left[\frac{T_p}{\eta_{T(W_c)}}\right]^{\beta_{T(W_c)}}} = R_0, \tag{52}$$

or

$$T_p = \eta_{T(W_c)}\left[\log_e \frac{1}{R_0}\right]^{\frac{1}{\beta_{T(W_c)}}}. \tag{53}$$

The **operational reliability** for any mission time T for the components with preventive replacement every T_p hours of cumulative operation, $\mathcal{R}(T)$, is [8]

$$\mathcal{R}(T) = [R(T_p)]^n R(\tau). \tag{54}$$

where

$$n = INT\left(\frac{T}{T_p}\right) = \text{integer part of } \left(\frac{T}{T_p}\right),$$

$$n \geq 0,$$

and

$$\tau = T - n \times T_p.$$

If the normal distribution is used for $T(W_c)$, then substituting Eq. (41) into Eq. (54) yields

$$\mathcal{R}(T) = \left\{\Phi\left[\frac{\mu_{T(W_c)} - T_p}{\sigma_{T(W_c)}}\right]\right\}^n \Phi\left[\frac{\mu_{T(W_c)} - \tau}{\sigma_{T(W_c)}}\right]. \tag{55}$$

If the Weibull distribution is used for $T(W_c)$, then substituting Eq. (42) into Eq. (54) yields

$$\mathcal{R}(T) = e^{-\left\{n\left[\frac{T_p}{\eta_{T(W_c)}}\right]^{\beta_{T(W_c)}} + \left[\frac{\tau}{\eta_{T(W_c)}}\right]^{\beta_{T(W_c)}}\right\}}. \tag{56}$$

273

Case 2 – Distributed Allowable Wear: Given the normally distributed allowable wear $W_c \sim N(\mu_{W_c}, \sigma_{W_c})$, the preventive replacement time T_p for a desired in-service component reliability $R(T_p) = R_0$ can be determined as follows:

If the normal distribution is used for $W(T_p)$, then

$$
\begin{aligned}
R(T_p) &= P[W_c - W(T_p) > 0], \\
&= \Phi\left[\frac{\mu_{W_c} - \mu_{W(T_p)}}{\sqrt{\sigma_{W_c}^2 + \sigma_{W(T_p)}^2}}\right], \\
&= R_0,
\end{aligned}
$$

and

$$
\frac{\mu_{W_c} - \mu_{W(T_p)}}{\sqrt{\sigma_{W_c}^2 + \sigma_{W(T_p)}^2}} = \Phi^{-1}(R_0), \tag{57}
$$

or

$$
\frac{\mu_{W_c} - \mu_{W(T_p)}}{\Phi^{-1}(R_0)}^2 = \sigma_{W_c}^2 + \sigma_{W(T_p)}^2. \tag{58}
$$

The value of T_p can be obtained by substituting Eq. (29) into Eq. (58) and solving for T_p. If the Weibull distribution is used for $W(T_p)$, we can not get the explicit solution for T_p. A trial-and-error procedure and Monte-Carlo simulation may be used together to find its numerical solution.

Preventive Replacement Scheduling for Minimum Cost: Given the allowable wear, W_c, failure replacement time and cost, t_f and C_f, respectively, and preventive replacement time and cost, t_p and C_p, respectively, the optimum preventive replacement age (cumulative in-service hours of operation) for the component, T_p, can be determined as follows:

The expected total cost per unit service time, $C(T_p)$, is [9; 10]

$$
C(T_p) = \frac{C_p R(T_p) + C_f[1 - R(T_p)]}{(T_p + t_p) R(T_p) + (M(T_p) + t_f)[1 - R(T_p)]}, \tag{59}
$$

where

$$
\begin{aligned}
M(T_p) &= \text{mean life (mean time to the allowable wear level } W_c) \text{ of the} \\
&\quad \text{component with preventive replacement at age } T_p, \\
&= \left[\int_0^{T_p} T_{W_c} f(T_{W_c}) dT_{W_c}\right] / [1 - R(T_p)],
\end{aligned}
$$

$$
f(T_{W_c}) = \text{pdf of the time to the allowable wear level } W_c,
$$

and

$$
R(T_p) = P[T_{W_c} > T_p] = \int_{T_p}^{\infty} f(T_{W_c}) \, dT_{W_c}.
$$

The optimum replacement age, T_p^*, is the one minimizing the expected total cost per unit time $C(T_p)$ given by Eq. (59). Therefore, T_p^* is the solution of the following optimization problem:

$$
\begin{aligned}
&Min \quad C(T_p) \\
&subject\ to \quad T_p > 0.
\end{aligned} \tag{60}
$$

If $T(W_c)$ is normally or Weibull distributed, we can get the numerical solution for T_p^* by a computer program.

274

TABLE 1– Wear versus operating time envelope data.

No.	Lower boundary		Upper boundary	
	Operating time, hr	Wear, in	Operating time, hr	Wear, in
1	0	0.0000	0	0.0000
2	1	0.0020	1	0.0027
3	2	0.0031	2	0.0039
4	3	0.0040	3	0.0046
5	4	0.0044	4	0.0051
6	5	0.0047	5	0.0054
7	10	0.0050	10	0.0058
8	15	0.0053	15	0.0060
9	20	0.0057	20	0.0063
10	25	0.0060	25	0.0068
11	30	0.0062	30	0.0070
12	31	0.0065	31	0.0073
13	32	0.0070	32	0.0079
14	33	0.0078	33	0.0087
15	34	0.0089	34	0.0100
16	35	0.0108	35	0.0128

Numerical Example: Given the observed wear versus operating time envelope data for hydrotreated fuel lubricated aircraft splines as listed in Table 1. Do the following:

1. Fit equations to these envelope data using Eq. (1).

2. Predict the wear reliabilities using the normal distribution assumption for the following cases:

   (a) Given the specified allowable wear $W_c = 0.0122$ in and the mission time $T_0 = 31$ hr.

   (b) Given the specified allowable wear $W_c = 0.0122$ in and the duty cycle time t of 31 ± 0.5 hr.

   (c) Given the normally distributed allowable wear $W_c \sim N(0.0122, 0.0001)$ in and the duty cycle time $T_0 = 31$ hr.

3. Determine the optimum preventive replacement time (age), T_p, for the following requirements:

   (a) Given the allowable wear $W_c = 0.0122$ in and the required component in-service reliability $R(T_p) = 0.9856$, using the Weibull distribution.

   (b) Given the normally distributed allowable wear $W_c \sim N(0.0122, 0.0001)$ in and the desired component in-service reliability $R(T_p) = 0.9856$, using the normal distribution.

   (c) Given the allowable wear, $W_c = 0.0122$ in, failure replacement time and cost, $t_f = 1$ hr and $C_f = \$1,000$, respectively, and the preventive replacement time and cost, $t_p = 0.5$ hr and $C_p = \$10$, respectively, using the Weibull distribution and minimum cost criterion.

275

Solutions to the Numerical Example:

1. Following the Steps 1 through 6, given earlier, yields the following: The lower boundary equations are

$$W_l(T) = \begin{cases} 0.002033 T^{0.5838}, & \text{for } 0 \le T < 4.21 \text{ hr,} \\ 6.2286 \times 10^{-5}(T - 4.21) + 0.004705, & \text{for } 4.21 \le T < 28.22 \text{ hr,} \\ 7.8150 \times 10^{-4}(T - 28.22)^{1.67065} + 0.0062, & \text{for } T \ge 28.22 \text{ hr,} \end{cases}$$

$$(61)$$

with the corresponding correlation coefficients of $\rho_0 = 0.9957$, $\rho_1 = 0.9968$ and $\rho_2 = 0.9957$.

The upper boundary equations are

$$W_u(T) = \begin{cases} 0.002746 T^{0.4622}, & \text{for } 0 \le T < 4.26 \text{ hr,} \\ 6.4571 \times 10^{-5}(T - 4.26) + 0.005366, & \text{for } 4.26 \le T < 28.93 \text{ hr,} \\ 8.5708 \times 10^{-4}(T - 28.93)^{2.5350} + 0.00696, & \text{for } T \ge 28.93 \text{ hr,} \end{cases}$$

$$(62)$$

with the corresponding correlation coefficients of $\rho_0 = 0.9962$, $\rho_1 = 0.9934$ and $\rho_2 = 0.9972$.

2. (a) The normal distribution parameters of wear at time $T_0 = 31$ hr can be determined by Eq. (29) in which $W_u(T_0)$ and $W_l(T_0)$ are calculated by substituting $T_0 = 31$ hr into Eqs. (61) and (62), respectively. The results are the following:

$$\begin{cases} W_l(T_0) = W_l(31 \text{ hr}) = 0.0105 \text{ in,} \\ W_u(T_0) = W_u(31 \text{ hr}) = 0.0124 \text{ in,} \end{cases}$$

and

$$\begin{cases} \mu_{W(T_0)} = \frac{0.0105 + 0.0124}{2} = 0.01145 \text{ in,} \\ \sigma_{W(T_0)} = \frac{0.0124 - 0.0105}{6} = 0.0003166 \text{ in.} \end{cases}$$

Then, from Eq. (38)

$$R(31 \text{ hr}) = \Phi\left(\frac{0.0122 - 0.01145}{0.0003166}\right) = \Phi(2.37) = 0.9911.$$

(b) Assume the duty cycle time t is normally distributed, then $\mu_t = 31$ hr and $\sigma_t = (1/6) = 0.1667$ hr assuming the duty cycle time tolerance (0.5) is $3\sigma_t$. The normal distribution parameters of time to the allowable wear $W_c = 0.0122$ in can be determined by Eq. (30) in which $T_u(W_c)$ and $T_l(W_c)$ are obtained by substituting $W = W_c = 0.0122$ in into Eqs. (61) and (62), and solving for T, respectively. The results are

$$\begin{cases} T_l(W_c) = T_l(0.0122 \text{ in}) = 30.9726 \text{ hr,} \\ T_u(W_c) = T_u(0.0122 \text{ in}) = 31.6074 \text{ hr,} \end{cases}$$

and

$$\begin{cases} \mu_{T(W_c)} = 31.29 \text{ hr,} \\ \sigma_{T(W_c)} = 0.1058 \text{ hr.} \end{cases}$$

Then, from Eq. (44)

$$R(t) = \Phi\left(\frac{31.29 - 31}{\sqrt{0.1058^2 + 0.1667^2}}\right) = \Phi(1.47) = 0.9292.$$

(c) From Case (a) we know that $\mu_{W(31\ hr)} = 0.01145$ in and $\sigma_{W(31\ hr)} = 0.0003166$ in. Then, applying Eq. (46) yields

$$R(31\ hr) = \Phi\left(\frac{0.0122 - 0.01145}{\sqrt{0.0003166^2 + 0.0001^2}}\right) = \Phi(2.26) = 0.9881.$$

3. (a) The Weibull distribution parameters of time to the allowable wear $W_c = 0.0122$ in are determined by Eq. (36) and are the following:

$$\begin{cases} \hat{\beta}_{T(W_c)} &= \frac{8.4952}{\log_e[31.6074/30.9726]} = 418.7235, \\ \hat{\eta}_{T(W_c)} &= 30.9726 \times 0.00135^{-\frac{1}{418.7235}} = 31.4652\ hr. \end{cases}$$

Then, from Eq. (53)

$$T_p = 31.4652\left(\log_e \frac{1}{0.9856}\right)^{1/418.7235} = 31.1487\ hr.$$

(b) From Eq. (51)

$$\begin{aligned} T_p &= 31.29 + 0.1058\Phi^{-1}(1 - 0.9856), \\ &= 31.29 + 0.1058\Phi^{-1}(0.0144), \\ &= 31.29 + 0.1058 \times (-2.1835), \\ &= 31.0590\ hr. \end{aligned}$$

(c) From Case (a) we know that $\hat{\beta}_{T(W_c)} = 418.7235$ and $\hat{\eta}_{T(W_c)} = 31.4652$ hr. Substituting the given data into Eq. (59) and solving Eq. (60), using a computer program, yields

$$T_p^* = 31.2903\ hr,$$

and

$$C(T_p^*) = \$779.0905/hr.$$

Conclusions: The results of this paper lead to the following conclusions:

1. The wear process is a stochastic process with random initial wear, random wear at a specified operating time, random time to a specified wear, random wear level at failure and random time to a failure wear level.

2. Typical wear process consists of three periods; i.e., break-in period, steady wear period and rapid wear (wear-out) period.

3. The wear process can be described by two distribution families; i.e, an wear distribution family at any operating time point and a distribution family of times to any wear level. These two distribution families form two wear-life envelopes which can be fitted to curvilinear-linear-curvilinear equations using the least-squares regression technique.

4. Distribution parameters can be obtained from the envelope data using the "3σ" theorem for the normal distribution and matching percentiles method for the Weibull distribution.

5. Wear reliabilities for fixed or distributed allowable wear, and fixed or distributed mission time can be predicted using the methodologies developed in this paper.

6. Preventive replacement time (age) for the specified in-service reliability or for minimum cost can be determined using the methods presented in this paper.

7. The methodologies presented in this paper can be applied to other failure modes exhibiting cumulative damage behaviors, such as metal fatigue, fatigue crack growth, corrosion, erosion, creep, deteriorating material properties in plastics with time, and so on.

References

1. Kececioglu, D. and Koharcheck, A., "Wear Reliability of Aircraft Splines", *Proceedings of 1977 Annual Reliability and Maintainability Symposium*, pp. 155-163, 1977.

2. Kececioglu, D. and Koharcheck, A., "Reliability of Aircraft Splines With or Without a Wear Induction Period", *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 105, pp. 163-170, April 1983.

3. Bogdanoff, J. L. and Kozin, F., *Probabilistic Models of Cumulative Damage*, John Wiley & Sons, Inc., 341 pp., 1985.

4. Kececioglu, D., *Reliability Engineering Handbook*, Prentice Hall, Inc., Englewood Cliffs, NJ 07632, Vol. 1, 688 pp., 1991.

5. Kececioglu, D., *Reliability and Life Testing Handbook*, Prentice Hall, Inc., Englewood Cliffs, NJ 07632, Vol. 1, 960 pp., 1993.

6. Kececioglu, D., "Reliability Analysis of Mechanical Components and Systems", *Nuclear Engineering and Design*, North-Holland Publishing Co., Vol. 19, No. 2, pp. 259-290, May 1972.

7. Kececioglu, D., "Fundamentals of Mechanical Reliability Theory and Applications to Vibroacoustic Failures", *Reliability Design for Vibroacoustic Environments, -Conference Proceedings by The American Society of Mechanical Engineers*, AMD, Vol. 9, pp. 1-37, Nov. 1974.

8. Kececioglu, D., *Maintainability Engineering Handbook*, The University of Arizona, 827 pp., 1992.

9. Jardine, A. K. S., *Maintenance, Replacement and Reliability*, Pitman, 199 pp., 1979.

10. Mou, Zhi-Zhong and Sun, Feng-Bin, "Multimode Replacement Policies for Large Scale Industrial Systems", *Proceedings of International Symposium on Reliability and Maintainability*, Tokyo, Japan, 1990.

DYNAMICS IN MONITORING GEAR FAULTS

Erkki Jantunen, MSc(Tech), Research Scientist
Antti Poikonen, BSc(Tech), Research Engineer
Technical Research Centre of Finland
Laboratory of Production Engineering
Espoo, Finland

Abstract: The study represents an attempt to simulate the vibration of a gearbox during accelerated wear tests in a laboratory. The natural frequencies and corresponding vibration modes of the shaft, gear and supporting structure are calculated with simplified finite element models. Based on results from oil analysis and vibration measurements, a very simple numerical formula describing wear on the gear teeth during the test runs is developed. Dynamic loads on gear teeth are calculated as a function of wear. These loads, together with vibration excitation from bearings and imbalance, are used in the calculation of the dynamic response. This calculation is performed on four occasions during the lifetime of the gearbox. The results obtained from dynamic response calculations are analysed with an FFT analyser, using the same methods of analysis as were used in the laboratory tests.

Introduction: An unexpected breakdown of machinery can cause a lot of damage. This has led to a growing need for effective and reliable condition monitoring methods. Today there are a great variety of methods available for monitoring rotating machinery. At the Technical Research Centre of Finland, the effectiveness and reliability of many of these methods have been studied in accelerated wear tests on a gearbox. These tests gave further information on the suitability of the methods tested for the condition monitoring of rotating machinery. After completing the tests, it was considered that further information could possibly be obtained by building a mathematical model of the test arrangement. The idea of this study is that, with a mathematical model, the tests can be simulated without unknown noise in the measured signals. Furthermore, it is possible with the mathematical model to study the simulated measuring signals at any of the nodes of the finite element model of the test arrangement and also in a stable situation at any chosen time during the entire lifetime of the gearbox.

Laboratory Tests: The wear and failure of a one-step gearbox (Santasalo 1C80) was studied in the laboratory by using vibration measurements and oil analyses. Two separate tests were run and during the tests the gear was overloaded by about 50% in order to accelerate wear (Kuoppala et al., 1991 and Aatola & Leskinen, 1990). In the first test, the gearbox ran for a total of 497 hours, until finally three teeth of the pinion broke at the base as a result of fatigue. In the second test, the gearbox ran for almost four times as long, i.e. 1945 hours, before similar failure.

279

The power of the gearbox was rated 17.9 kW, and the number of teeth of the pinion and gear were 23 and 101, respectively, with a gear ratio of 4.3913. The electric motor (VEMKMER 225 54 AC P) ran at a constant speed of 1500 rpm (25 Hz) and consequently the speed of the output gear was about 342 rpm (5.7 Hz). The power was transmitted to a pneumatically controlled mechanical disc brake (Aatola & Leskinen, 1990).

Vibration signals were recorded at five measuring points on the gearbox and also on the electric motor and the brake. The following analyses of vibration signals were used: spectrum and cepstrum analysis (using time averaging synchronized with the running speed of both input and output shafts, and also without synchronization using spectrum averaging); acoustic emission; statistical analysis (rms, peak and kurtosis) and synchronized time domain signal analysis. In the first test, the oil analyses consisted of automatic particle counting, ferrography and spectrometric oil analysis, and in the second test, wear particle analysis using an on-line wear particle sensor constructed at the Technical Research Centre of Finland (Kuoppala et al., 1991).

In both of the laboratory tests, cepstrum analysis (0 - 500 Hz) was able to predict the upcoming failure by monitoring components corresponding to the speed of the shafts. The synchronized cepstrum analysis was more sensitive than the ordinary cepstrum analysis. In the first test, the cepstral component of the input shaft provided an indication of failure more than four hours before failure. In spectrum analysis, the spectral sidebands around the gear mesh frequency were rather unstable and showed the upcoming failure about ten minutes before it occurred. The spectral running speed component of the input and output shafts did not give an indication of the upcoming failure. Synchronized time domain signal analysis showed the upcoming failure clearly one hour before failure. All the statistical parameters, the rms, peak and kurtosis values showed minor changes about one hour before the failure. Acoustic emission did not give a reliable indication (Aatola & Leskinen, 1990). The correlation between the three different oil analysis methods was good during the tests (Kuoppala, et al., 1991). Wear particle analysis, with the on-line sensor which was installed in the second test, is used later in this study for the definition of wear and dynamic loads.

Finite Element Model: A simple finite element model was made of the entire laboratory test arrangement, consisting of a shaft, gearbox and supporting structure, using a graphical Patran modelling package. The total number of degrees of freedom in the model was 10236 (2145 elements, 2443 nodes). The whole model is shown in figure 1.

The shaft was modelled very roughly using beam elements. This simplified approach was chosen for two reasons. First, the size of the model was to be kept small. The second reason was that there were no measured values of torque available from the laboratory tests in a suitable form for comparison. The pinion and the gear were simply modelled with two beam elements attached at right angles to each of the shafts in the mesh. The lengths and cross-sections of these beams were chosen to represent the gear ratio, inertias and tooth flexibilities (Lees & Pandley, 1980). The shaft model was connected through translational degrees of freedom to the rest of the model at the bearings.

The gear casing was modelled with isoparametric linear solid elements. The idea originally was to model the gear casing geometrically fairly precisely so that the local

280
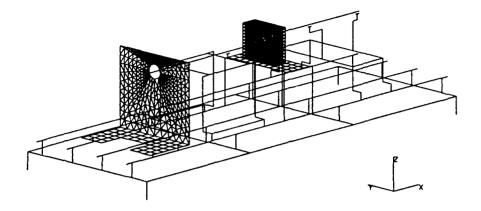
Fig. 1. Finite element model of a laboratory test arrangement consisting of a shaft, a gearbox and a supporting structure.

behaviour of the structure at the measuring points could be calculated accurately. Unfortunately, this goal could not be achieved with reasonable effort, and consequently a very simplified model of the gear casing had to be used. The supporting structure was modelled with beam and shell elements. This part of the model was considered the most unimportant part of the whole model, so a very coarse element mesh was used. The gear casing was connected to the supporting structure through common nodes. For the testing of the whole calculation and process of analysis and especially the dynamic loads, an additional local dummy model with only 30 degrees of freedom was also developed.

Natural Frequencies: Because the gear mesh frequency is 575 Hz, it was considered necessary to verify the higher natural frequencies. Therefore impact hammer tests were included in this study. Mechanical accelerance was measured in a broad frequency range from 0.2 Hz to 1 kHz. In these measurements, coherence was typically over 0.95 up to 600 - 800 Hz, so the most interesting frequency range could be covered. In these tests, the supporting structures under the brake, gear casing and electric motor were excitated and the response was measured separately. All of the structures were measured in the longitudinal and transversal direction. The gear casing was also tested in the vertical direction. A number of natural frequencies were found in the impact tests, but none of these were close to the main excitation frequencies. The most dominant natural frequencies found in the impact hammer tests are shown in table 1 (together with the corresponding calculated values).

It was not expected that the calculated values would correlate exactly with the measured ones, because of the coarse modelling technique and also because the measuring conditions did not correspond exactly to the modelled situation. The natural frequencies of the finite element model were calculated with the Abaqus program package. The frequency analysis was limited to 700 Hz, which was considered to be high enough over the gear mesh frequency (575 Hz). A total of 100 eigenvalues were found. This was considered far too many for the dynamic response analysis, especially since most of them were local modes and as such irrelevant.

281

Table 1. Comparison of calculated and measured natural frequencies.

	Calculated, FEM model			Measured, Impact hammer test		
Coordinate axis	X	Y	Z	X	Y	Z
Mode number FEM model	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]
2		61		60		
3		90			98	
4	111	111	111	113		
5	116	116	116	115	116	
12	356		356			387
20	741			740		

In order to reduce the number of eigenvalues and the time needed for the dynamic response analysis, superelement techniques were adopted. The number of degrees of freedom was limited to 108, and the 25 lowest eigenvalues were calculated. In table I, the frequency of some of these modes is compared with the natural frequencies found in the impact hammer tests. The correlation between measured and calculated values seems to be rather good. Unfortunately, it was not possible to compare the actual natural modes because of the simple testing arrangement in the impact hammer tests, and consequently the results shown in table I might also give an excessively optimistic view. Based on the results from impact tests, it was considered important to ensure that no natural frequencies would exactly match any of the frequencies of the dynamic loads, because that would have led to resonance in the calculation of dynamic respon-se. The calculated natural frequencies fulfilled this condition.

Development Of Dynamic Loads: The wear of the gear teeth is a complicated phenomenon and is a function of a number of parameters, such as temperature, pressure, the hardness of the sliding surface, sliding velocity etc. (Holmberg, 1991). In this study, a very simplified approach, which is a numerical method and not physically explained, was adopted. Based on a number of studies, Onsøyen (1991) has summarized a simple model for the wear depth

$$h(t) = h_0 + h't$$

where $h(t)$ is the wear depth, t is the time, h_0 is the contribution from running-in and h' is the wear rate (the increase in wear depth per unit of time). The time to failure is the time t_c until $h(t)$ reaches critical wear depth h_c. It was assumed that the wear progression during the laboratory tests had been of a progressive type (Onsøyen, 1991) so that the wear behaviour at the beginning can be described as mild wear and at the end as severe wear (Holmberg, 1991). To fulfil this assumption, a simplified numerical expression for the wear rate was chosen

$$h'(t) = A * t_c / (t_c - t)$$

where A is a coefficient which does not vary as a function of time. For simplicity, running-in wear is not accounted for in the above expression. By integrating the above formula, a numerical expression for the wear depth was developed

$$h(t) = - A * t_c * Ln(1 - t/t_c)$$

From the second laboratory test results, it was known that the failure first took place at or f the pinion teeth and that t_c for that tooth was 1945 hours. During the second laboratory test, an on-line wear particle sensor was used for three time periods (Kuoppala et al., 1991). At the beginning of these periods, the collecting rapidity of

282

wear particles was 4.0 / cycle 1 (t = 0 hours corresponding to $Ln(t_c/(t_c - t)) = 0$), 6.3 / cycle 2 ($Ln(t_c/(t_c - t)) = 1.4$) and 15.0 / cycle 3 ($Ln(t_c/(t_c - t)) = 3.7$). Based on these recordings and the formula for the wear rate, t_c was numerically solved for the other pinion teeth, and turned out to be 1.41 times longer. It was then assumed that t_c for the gear is the gear ratio times t_c for the pinion.

Spotts (1984) gives a simple formula for estimating dynamic loads on gear teeth caused by manufacturing errors

$$F_{dm} = 2*e_t*(k*m_e)^{1/2}/t_a$$

where e_t is the total error for a tooth pair, k is the tooth pair stiffness, m_e is the effective mass for a gear pair and t_a is the tooth error application time. In this study, it was assumed that the manufacturing errors and wear depth have a similar effect on the dynamic loads on gear teeth; thus, the total dynamic load was

$$F_{dt} = F_{dm} + F_{dw}$$

where F_{dw} is the dynamic load caused by wear, i.e. a linear function of wear depth, which was used instead of e_t in the above formula when F_{dw} was estimated. For the pinion and gear, the manufacturing errors were known approximately from the manufacturing tolerances, and all other parameters needed for the calculation of F_{dm} were a function of the pinion and gear geometry. For the calculation of F_{dw}, one further assumption had to be made; i.e., it was assumed that in the mild wear region the vibration level at the gear mesh frequency was an indicator of F_{dw}. Based on the vibration measurement results, it was assumed that for all other teeth, except the one that would cause the failure of the gearbox, F_{dt} at the end would be 2 times F_{dt} at the beginning. Figure 2 shows, as a function of $Ln(t_c/(t_c - t))$, the normalized sum of dynamic loads on the gear teeth. At lower values of $Ln(t_c/(t_c - t))$, the teeth of the pinion which were assumed to have a longer lifetime exercise a considerable influence on the increase of the dynamic loads, but at higher values of $Ln(t_c/(t_c - t))$, the linear effect of the individual tooth which was assumed to cause the failure is observed.
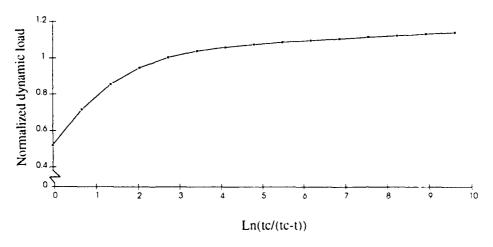
Fig. 2. Normalized sum of dynamic loads on gear teeth.

The assumptions described above are rather radical, but after the dynamic analysis (Jantunen & Poikonen, 1992) described later in this paper, their suitability was further compared in detail against the vibration data from the second gearbox test. On the basis of earlier analyses it was known that the overall acceleration level (RMS value)

283

does not show upcoming failure clearly, so in this new analysis it was assumed to indicate other changes during the tests, i.e., variations in power or the dynamic behaviour of the brake. Figure 3 shows, as a function of $Ln(t_c/(t_c - t))$, the peak-to-peak value of the synchronized time domain acceleration signal (averaged in time domain, number of samples 100) which before normalization has been divided by the square root of the overall acceleration level (between 10 and 1000 Hz) and the cepstral running speed component of the input shaft (quefrency 40 ms, synchronized with the running speed of the input shaft, averaged in time domain, number of samples 100) which before normalization has been divided by the acceleration overall level raised to the power of 0.25, i.e., it has been assumed that the peak-to-peak value is more sensitive to changes in the measuring condition. In this analysis, linearity in the growth of the peak-to-peak value and cepstral component at higher values of $Ln(t_c/(t_c - t))$ can be observed. From the assumptions given above, it also follows that no correlation could be expected at lower values of $Ln(t_c/(t_c - t))$ between the total load shown in Figure 2 and the analysed vibration signals shown in Figure 3. It should also be noted that this kind of scaling of the cepstral component and peak-to-peak value of acceleration with the overall vibration level was considered possible because the running speed of the shafts had been nearly constant throughout the test, so it could be anticipated that the influence of the natural modes had not varied dramatically.

$$Ln(tc/(tc-t))$$
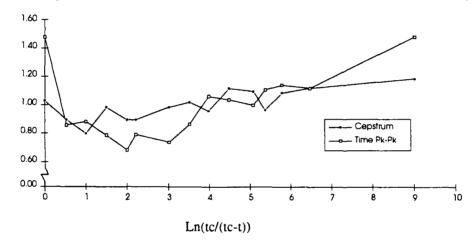
Fig. 3. Normalized cepstral running speed component of the input shaft and normalized peak-to-peak value of synchronized time domain acceleration signal.
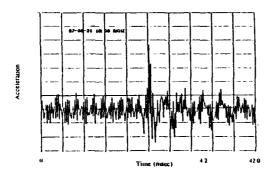
In the analysis, it was assumed that only the gear teeth suffer from wear. All other dynamic loads, i.e., imbalance and bearing forces, were introduced as constants. This assumption was made because no measured values were available, and on the other hand it was considered that the variation in the imbalance is very small within time and that the bearings have a considerably longer lifetime than the gear teeth. The imbalance of the electric motor and the input shaft, and similarly the imbalance of the brake and the output shaft connected to it, were introduced into the model at the bearings. The corresponding frequencies were 25 Hz and 5.18 Hz. It was also assumed that there are some faults in all the bearings and all their components. In reality this assumption is not true, but it served as a means of specifying the frequencies of artificial noise. The corresponding excitation frequencies were calculated with well-known formulae (e.g. Springer, 1988)

284

$$BPFO = Nb/2^*S^*(1 - Bd/Pd^*\cos \phi)$$
$$BPFI = Nb/2^*S^*(1 + Bd/Pd^*\cos \phi)$$
$$FTF = S/2^*(1 - Bd/Pd^*\cos \phi)$$
$$BSF = Pd/(2^*Bd)^*S^*(1 - (Bd/Pd)^{2*}(\cos \phi^2))$$

where Nb is the number of rolling elements, Bd is the diameter of the rolling elements, Pd is the pitch diameter, ϕ is the contact angle and S is the speed of rotation. The calculated frequencies were BPFO (ball pass frequency, outer race), which indicates a fault in the outer race, BPFI (ball pass frequency, inner race), which indicates a fault in inner race and BSF (ball spin frequency), which indicates a fault in a rolling element. As no measured data on the size of these forces were available, their size was chosen in relation to the dynamic loads on the gear teeth.

Dynamic Response: The dynamic response of the FEM model, constructed using superelement techniques with a reduced number of degrees of freedom, as well the dummy model were calculated with the above described dynamic loads and using the so-called step-by-step integration method. The response was calculated during four phases of the laboratory tests, namely, at the beginning of the tests, after 50% of the hours of running (halfway, corresponding to $Ln(t_c/(t_c - t)) = 0.69$), and at 2.6% of lifetime remaining ($Ln(t_c/(t_c - t)) = 3.6$) and 0.2% of lifetime remaining before failure ($Ln(t_c/(t_c - t)) = 6.2$). The length of the phases was limited to 1.024 seconds and the time step was 0.390625 milliseconds. Calculated acceleration was tabulated as a function of time at a node corresponding to a measuring point in the laboratory tests. The tabulated results were transferred to a PC and from there using a DA-card, to a FFT spectrum analyser. For practical reasons, this step was performed ten times more slowly than in reality. Similar methods of analysis were used as had been used with measured signals in the laboratory tests.

The time domain signals from laboratory tests and mathematical simulation (dummy model) are compared when there was 0.2% of lifetime remaining before failure, in Figures 4a and 4b. As can be seen, the time domain signals correlate fairly well with each other and the effect of the one tooth which eventually broke first is seen.
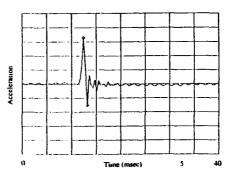
Fig. 4a. Measured time domain signal.   Fig. 4b. Simulated time domain signal.

The spectrums calculated (dummy model) with 2.6% and 0.2% of lifetime remaining before failure are shown in a broad frequency range in Figures 5a and 5b. In the spectrums, the somewhat irrational behaviour of spectral sideband components and

285

the rather small changes in the running speed component of the input shaft and the gear mesh component can be seen. These trends are similar with reported laboratory test results (Aatola & Leskinen, 1990 and Kerkkänen & Kuoppala, 1990) as described earlier in this paper.
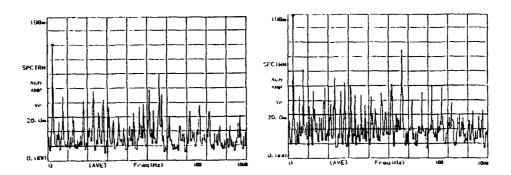
Fig. 5a. Spectrum 2.6% of lifetime remaining. Fig. 5b. Spectrum 0.2% of lifetime remaining.

The finite element model with a reduced number of degrees of freedom, constructed using a superelement technique, was not capable of showing the higher frequency range, i.e. frequencies over 250 Hz. Apparently the chosen approach in reducing the number of degrees of freedom was far too radical, and vital local modes between the loads and measuring points/nodes were lost. The spectrum at measuring point 3 of the laboratory tests in the lower frequency range corresponding to the beginning of the tests, with 2.6% and 0.2% of lifetime remaining before failure, are shown in Figures 6a, 6b and 6c. From these spectrums, too, it is rather difficult to judge whether or not a failure is imminent. However, the increase in the running speed component of the input shaft (25 Hz) is about 3.5 dB between 2.6% and 0.2% of lifetime remaining before failure. This does not correlate with reported measurement results (Aatola & Leskinen, 1990 and Kerkkänen & Kuoppala, 1990).

Figures 7a, 7b and 7c show the results from cepstrum analysis at measuring point 3 of the laboratory tests (calculated in the frequency range from 0 to 1 kHz) corresponding to the beginning of the tests, with 2.6% and 0.2% of lifetime remaining before failure. Unfortunately, due to the restrictions of the analyser used, the cepstrum is shown in the same way as a time domain signal, which is not the normal way of presenting the cepstrum. Although the spectrum does not show any noticeable change when only 2.6% of lifetime is remaining before failure, the corresponding cepstrum differs markedly from the cepstrum of the beginning phase. With between 2.6% and 0.2% of lifetime remaining before failure, a further change can be noticed. This trend, if the absolute values are not studied, is similar to that found in the analyses of measured data (Aatola & Leskinen, 1990 and Kerkkänen & Kuoppala, 1990). The effectiveness of cepstrum analysis is based on its ability to detect periodicity in the spectrum, e.g. families of harmonics and uniformly spaced sidebands, while it is insensitive to the transmission path of the vibration signal from its origin to the external measuring point (Randall & Hee, 1981), i.e., it separates noise from other sources quite well. In the mathematical model, the increase in vibration due to the higher dynamic loads on gear teeth is scattered to the harmonic components of the running speed of the input
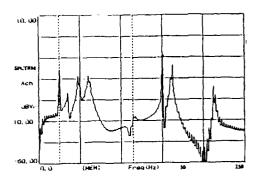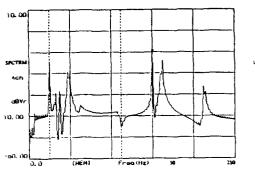
Fig. 6a. Spectrum at the beginning. Fig. 7a. Cepstrum at the beginning.
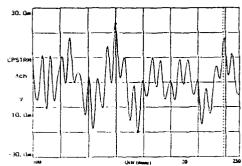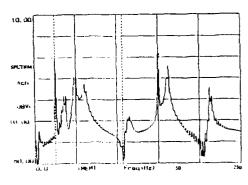
Fig. 6b. Spectrum 2.6% of lifetime remaining. Fig. 7b. Cepstrum 2.6% of lifetime remaining.
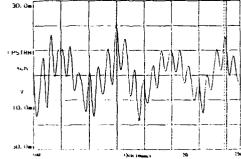
Fig. 6c. Spectrum 0.2% of lifetime remaining. Fig. 7c. Cepstrum 0.2% of lifetime remaining.

287

shaft, including the sidebands of the gear mesh frequency, rather randomly and is therefore detected with cepstrum analysis.

From the definition of the dynamic loads in the mathematical model, it follows that the overall vibration level in a broad frequency range is a poor indicator of an upcoming failure, since the increase in loads due to wear as a function of time is relatively small compared to the sum of all of the loads introduced into the model. The mathematical model also points out how the indication with time domain signal analysis and cepstrum analysis is a function of the wear mechanism. For example, if it is assumed that the wear of one individual tooth would not differ so much from that of the other teeth, the indication would not be as clear.

Conclusion: Laboratory tests with a gearbox were simulated with coarse finite element models and using a simple numerical expression for wear based on the results from an on-line wear particle sensor. The results from dynamic response calculations were analysed in a manner similar to that used in obtaining results from laboratory tests. In spite of the great simplification of the mathematical model, by comparison with the actual tests, in many respect it reveals a correlation between trends of the analysed results. The mathematical model helps us understand why certain methods of analysis, i.e., cepstrum analysis and synchronized time domain signal analysis, are better tools for giving an indication of upcoming failure than the other methods tested are.

References:
Aatola, S. & Leskinen, R. (1990). Cepstrum analysis predicts gearbox failure. *Noise Control* Eng. J., 2, pp. 53-59.
Holmberg, K. (1991). Tribological bases for accelerated testing. In *Operational reliability and systematic maintenance*. Ed. K.Holmberg & A.Folkeson. Elsevier science, London, pp. 31-50.
Jantunen, E. & Poikonen, A. (1992). Dynamics in condition monitoring, a case study. Structural analysis 1992. *Technical Research Centre of Finland, VTT Symposium 128*, pp. 125-138.
Kerkkänen, K. & Kuoppala, R. (1990). Methods of machine failure diagnostics (Koneiden vikadiagnostisoinnin menetelmiä) (in Finnish). *Technical Research Centre of Finland, Research Reports*, No. 700.
Kuoppala, R.J., Jantunen, E.O. & Enwald, P.A. (1991). Condition monitoring methods for rotating machinery. In *Operational reliability and systematic maintenance*. Ed. K.Holmberg & A.Folkeson. Elsevier science, London, pp. 175-198.
Lees, A.W. & Pandley, P.C. (1980). Vibration spectra from gear drives. I Mech E C265/80 pp. 103-108.
Onsøyen, E. (1991). Accelerated testing of components exposed to wear. In *Operational reliability and systematic maintenance*. Ed. K.Holmberg & A.Folkeson. Elsevier science, London, pp. 51-77.
Randall, R.B. & Hee, J. (1981). Cepstrum analysis. *Brüel & Kjær Technical Review*, 3, pp. 3-40.
Spotts, M.F. (1984). Estimating dynamic loads on gear teeth. *Machine design*, July 12, pp. 84-88.
Springer, C.W. (1988). Spectral Analysis of double-row antifriction bearings. *Vibrations*, Vol. 4 No. 1 March, pp. 16-17.

FATIGUE RATCHET FAILURES IN PIPING SYSTEMS

R. J. Scavuzzo and P. C. Lam

The University of Akron
Akron, Ohio 44325-3904

Abstract: The mechanism of ratcheting in dynamically loaded straight piping and piping elbows is presented. In addition, test data is reviewed for pipe and elbows.

Simple equations developed by Edmunds and Beer and by Beaney can be used to bound ratcheting strains in straight pipe. These equations are reviewed. Results of more accurate elastic-plastic finite element analyses are also presented. Stress distributions in elbows are too complicated to use formulas based on a plane stress analysis to predict ratcheting strains. In this case, the only analytical approach known is elastic-plastic finite element analysis.

Introduction: Allowable stresses in pressurized piping subjected to shock or other dynamic loads are usually based on the yield strength or some multiple of the yield strength. These criteria, that lead to very safe design, ignore the true modes of failure. Resulting designs of pressurized piping systems for nuclear power plants can withstand seismic loads an order of magnitude higher then allowed by the ASME Boiler and Pressure Vessel Code. This conservatism is introduced because the Code does not address the correct mode of failure. Pressurized piping systems fail by fatigue or fatigue-ratcheting and not static collapse [1-4]. Large amounts of kinetic energy developed in piping systems can be absorbed by plastic cycling and thus create effective viscous damping of over 20% [1,2].

There are no accurate closed-form solutions of incremental plastic ratcheting of pressurized piping caused by axial bending from seismic loads that develop stresses into the plastic range. However, approximate solutions of the incremental plastic strains caused by ratcheting have been proposed by Edmunds and Beers [5], Beaney [6-9] and Miller [10]. Experimental work on straight pipe has been conducted by the EPRI [3,4], The University of Akron [11,12,13] and by Beaney in the United Kingdom [2,10] as well as others.

The phenomenon of ratcheting can be easily explained by considering a thin pipe wall to be a flat plate with a hoop stress caused primarily by the pressure loading in one direction and an oscillating axial stress perpendicular to the hoop stress as shown in Figures 1 and 2. This model was used by Edmunds and Beer [5]. Yielding was predicted based on the Tresca criteria. The oscillating axial stress is caused by a combination of the pressure loading and bending and is assumed to exceed the hoop stress in magnitude. The effective stress acting on the pipe must exceed the yield point for ratcheting to occur. First, assume that both the hoop stress and axial stress are tensile. If the yield point is exceeded, plastic flow will occur in the axial-radial plane. The pipe wall will become thinner as the axial direction grows (Fig. 1). Then the axial stress becomes negative from the bending as shown on Figure 2. In this case the plastic flow is in the axial-hoop plane. The hoop direction decreased as the axial direction increases. The overall effect in one cycle is that the pipe wall becomes thinner and the hoop direction increases. The plastic strain in the axial direction in straight pipe oscillates and does not ratchet. Using the Edmunds-Beer model, the calculated hoop strain can be plotted against the axial strain (Fig. 3). Ratcheting of the hoop strain during the compression cycle can be observed. There is no plastic flow during the tensile portion of the cycle. Beaney was the first researcher to use this type of plot to present ratcheting experimental data (Fig. 4) [8]. It should be observed that on Figure 4 the axial strain oscillation increases in magnitude as the test proceeded.

Ratcheting of Straight Pipes: A series of dynamic and static tests on straight pipe were conducted at the University of Akron [11-15]. In each case the pipe was pressurized and subjected to axial bending moments. The axial loading simulates dynamic bending from seismic or shock loading. The static test was a displacement controlled test using a four point loading as shown on Figure 5. Typical of the data on straight pipe is shown on Figures 6 and 7. It should be noted that the axial strains oscillate about zero unless the magnitude of the cyclic displacement is increased to a very large value. With this loading, there is some axial ratcheting as well as hoop ratcheting. Hoop ratcheting still dominates. Initially, the hoop ratcheting is very high. However, after 15 to 20 cycles the ratcheting reaches a constant rate. As expected, if the magnitude of the cyclic displacement increases, the magnitude of the ratcheting increases. Also, the rate of ratcheting is dependent on the internal pressure. Both 304SS and carbon steel pipes were tested. Ratcheting was similar for the two materials.

In the dynamic tests conducted at the University of Akron, axial bending was developed by inertia from the pipe and concentrated weights fixed to the pipe. Initial hoop ratcheting was followed by cyclic plasticity. After about 20

290

cycles, ratcheting stopped. There was not enough power in the system to continue the ratcheting. Furthermore, once a specimen was ratcheted in a run, only oscillation with the strains at or a little above the yield point could be obtained. There was no further ratcheting on subsequent runs [14].

In analytical work at the University of Akron, the nonlinear finite element code, ABAQUS [16], was used to study ratcheting of straight pressurized pipe subjected to cyclic bending loads. For elastic-plastic time dependent loading, the specified stress-strain curve must be bilinear. In both cases the elastic modulus is 28.5 Mpsi; the tangent modulus, E_t, was varied from zero to 5.5 Mpsi. Two values are shown on Figure 8: a perfectly plastic material with E_t equal to zero and with E_t equal to 500,000 psi. Typical results are shown on Figure 8 for the test geometry shown on Figure 5. In these analyses, the decrease in the rate of ratcheting to a uniform value can be observed. Also, the dependence of the ratcheting on the tangent modulus is evident. If the analysis allowed a more exact specification of material properties, more exact analytical comparisons with data could be developed. However, as discussed below, elastic-plastic finite element analyses are the most accurate analytical technique [15].

In the finite element analysis, the ELBOW 31 element in ABAQUS was specified. However, the default values of 5 integration points in the thickness direction and 16 integration points in the axial direction were increased to 11 and 33, respectively, in order to obtain the required accuracy. Results of the study on the accuracy of the elastic-plastic bending of beams and pipes is presented in Reference [13].

Approximate formulas that predict ratcheting strains in piping have been proposed by Miller, Edmunds and Beer and Beaney. The Edmunds-Beer and Beaney formulas are presented as equations (1) and (2), respectively:

$$d\epsilon/dN = 3\sigma_h/(2\sigma_y - \sigma_h)[2\epsilon_b - (2\sigma_y-\sigma_h)/E] \qquad [1]$$

and

$$d\epsilon/dN = 6\sigma_h/E(2\sigma_y - \sigma_h)[\sigma_a+\sigma_h/2)-\sigma_y] \qquad [2]$$

where
$$d\epsilon/dN = \text{ratchet strain per cycle}$$
$$\sigma_h = \text{hoop stress}$$
$$\sigma_y = \text{yield point}$$
$$\epsilon_b = \text{cyclic axial strain from bending}$$
$$E = \text{Young's modulus}$$
$$\sigma_a = \text{axial stress amplitude from bending}$$

Comparison of Analytical and Experiment Results: A comparison of straight pipe and the Edmunds-Beer and Beaney approximate

291

formulas is made on Figure 9. As seen on this graph, both formulas bound the measurements. However, Beaney's equation is more accurate and, therefore, is recommended.

Results from the displacement-controlled finite element model were compared with static test data in Table 1 [11]. The initial measured ratcheting hoop strain on the first cycle is about 1000 μin/in for the stainless steel specimen and 650 μin/in for the carbon steel specimen. Measured steady state ratcheting strain is obtained after about 30 cycles of the fixed displacement input. As indicated in Table 2, calculated incremental hoop strains based on a tangent modulus of 5,500,000 psi underestimate measured values. When the smaller tangent modulus of 500,000 psi is specified, incremental hoop ratcheting is overestimated.

Table 1

Comparison of Calculated Ratcheting Strain with Static Tests on Straight Pipe

FINITE ELEMENT ANALYSIS		TEST	
E_t 5.5x10^6	E_t 0.5x10^6	304 SS	Carbon Steel
500μ/cycle*	1400μ/cycle*	948μ/cycle*	650μ/cycle*
$\approx$10μ/cycle**	400μ/cycle**	17μ/cycle**	47μ/cycle**

Ratcheting strain is presented in micro-inches/inch per cycle

* First Cycle
** Steady State Value

Elbow Analysis: In-plane cyclic loading in the plastic range of pipe elbows was recently conducted at the Institut fur Stahlbau und Werkstoffmechanik in Darmstadt, Germany [17-20]. Several elbows were tested under different loading conditions. In general, the tests differed in the applied load history and internal pressure.

In-plane quasi-static cyclic loading was applied to a 90 degree pipe elbow using the test setup shown in Figures 10 and 11. The steel elbow tested was a St 35.2 which has an outside diameter of 219.1 mm (6.625"), a thickness of 6.3 mm (0.25"), and a radius of curvature of 305 mm (12"). The elbow was welded to a 400 mm straight pipe run on each side. Flanges were welded to the ends of the straight pipe runs and were bolted to the rigid test frame. Displacement controlled static load cycles were applied using a hydraulically driven actuator. The initial displacement amplitude was 5 mm (0.2") for the first two cycles and was increased by 5 mm every 2 cycles up to 50 mm as shown in Figure 12. The test specimen

292

(elbow and straight pipe runs) was pressurized with an
internal pressure of 1.2 KN/cm^2.

Strain gages were mounted at several locations in the middle
of the elbow as shown in Figure 11. The gages were placed to
measure the axial strains at 0, 30, 60, 90, 120, 180, and 270
degrees (0 degree is the intrados, gage 1). Hoop strains
were measured at 90 and 270 degrees only (the maximum stress
location for in-plane loading). All measurements were made
on the elbow outside surface. Strain data was automatically
stored in a digitized form and were obtained from the
investigator for comparisons with the finite element analysis
presented in this paper.

Finite Element Model Description: The elbow element ELBOW31
of the ABAQUS finite element program was also used to model
the elbow test specimen which included the elbow and attached
straight pipes. The rigid test frame used to apply the
bending moment was modeled by beam elements B31. The finite
element mesh is shown in Figure 13. These elements model
ovalization of the cross-section, warping, and ovalization
gradient along the elbow. The elements use linear polynomial
interpolation along their lengths together with Fourier
interpolation around the pipe for all motion relative to the
pipe axis. The number of Fourier terms in the series (called
Fourier ovalization modes) is limited to a maximum value of
6. The elements have one integration section along their
length with integration points around the pipe and through
the thickness.

A 5-7-5 mesh (5 elbow elements in each of the straight pipes
and 7 elements in the elbow) modeled the test specimen. Beam
type elements B31 (10 on each side) were used to model the
moment arm (rigid test frame). Thirty six integration points
in the circumferential direction and eleven integration
points through the thickness were specified for all elbow
elements in this study. Where the test specimen meets the
test frame at the flange, it was assumed that the flange was
stiff enough so that all cross-sectional deformation
(ovalization and warping) were restrained. At one end (node
1 in Figure 13) only rotation around the 3-axis was allowed
while at the other end (node 38 in Figure 4), where the
cyclic displacement is applied, only rotation around the 3-
axis and translation along the 1-axis were allowed. The
formulation for the internal pressure loading for the ELBOW31
element includes the hoop stress only and omits the axial
pressure stress. However the axial pressure stress was
introduced in the model by applying an equivalent axial force
at node 38.

The metal plasticity model applied in this study assumes a
Von Mises yield surface with associated plastic flow, and
kinematic hardening. The kinematic hardening yield surface
radius is constant but moves in the stress space during

straining and thus models the Bauschinger effect associated
with strain reversals. The kinematic hardening model
available in ABAQUS is a Prager-Ziegler model with uniaxial
response modeled by a bilinear elastic-plastic material. An
elastic modulus of 200,000 MPa, a yield point of 221 MPa
which is less than the yield strength of 290 MPa and a
tangent modulus of 1724 MPa were assumed in this study.
Furthermore, the straight pipes and the elbow were assumed to
have the same material properties.

Comparison of Finite Element Analysis with Elbow Test Data:
Comparisons of the ABAQUS finite element results with the
German test data are shown in Figures 14 - 19. On these
figures the predicted and measured strains versus cycles are
plotted. Strain gauge failures occured in the hoop
measurements shown on Figures 18 and 19 and the measurements
became a straight line after gage failure. Test data were
available at the gage locations shown in Figure 11. Axial
strains were measured at 0, 30, 60, 90, 120, 180, and 270
degrees while hoop strain was measured at 90 and 270 degrees
only. All strain measurements were made on the outside
surface in the middle of the elbow.

During the first few cycles where the strains remained mainly
elastic the ABAQUS results are in good agreement with the
test data. After the initial few cycles the applied
displacements caused strains in the plastic range and
ratcheting started at all gage locations in both test and
analysis. However, the measured and predicted axial
ratcheting was insignificant when compared to hoop ratcheting
at 90 and 270 degrees (Figures 18 and 19). Axial ratcheting
was not all in the same direction in either test or analysis
(Figures 14 - 17). Axial ratcheting at 30 and 60 degrees was
negative (Figures 15 and 16) while positive axial ratcheting
occurred at all other locations for which test data was
available. Where the axial ratcheting was negative, the
predicted ratcheting was even more negative, and where it was
positive the predicted ratcheting was more positive. The
degree to which the axial ratcheting was over predicted
(positive or negative) depends on the location around the
elbow. At 90 and 270 degrees (the maximum stress location
for in-plane loading, Figures 18 and 19) significant hoop
ratcheting occurred. Ratcheting predictions at these
locations by the finite element program were in good
agreement with the test data up to about the tenth cycle
after which conservative predictions were calculated.
Locations 90 and 270 degrees are the same direction because
of load symmetry for in-plane loading. However, the measured
hoop ratcheting was higher at the 90 degrees than at 270
degrees. The difference could be due to a combination of
several reasons which include out of circularity, nonuniform
elbow thickness and errors in strain gage placement.

The deviation of the finite element results from the test
data in the plastic range could be due to shortcomings in the
kinematic hardening model which is based on a bilinear

294

elastic plastic material defined by the elastic modulus, yield stress, and the tangent modulus. The assumed tangent modulus significantly influences the rate of ratcheting in elbows as shown in Reference [17]. Also, the exact stress-strain curve was not measured and approximate values for the yield stress and tangent modulus were specified. However, the analysis results support the applicability of the ABAQUS finite element program in the predictions of ratcheting in elbows.

Conclusions: By specifying elbow elements in the ABAQUS finite element program and the kinematic hardening rule for pressurized pipe subjected to axial bending, ratcheting of the hoop strain was calculated in both straight pipe and elbows.

Both test results and finite element analyses agree that ratcheting is influenced by the material stress strain curve and the loading history.

The rate of ratcheting depends significantly on the magnitude of the internal pressure and tangent modulus of the bilinear material.

The measured and calculated rate of ratcheting decreases with cycles in the straight pipe. At times there is shakedown and incremental hoop strain decreases to zero. These trends were observed in both static and dynamic ratcheting tests and finite element analyses of straight pressurized pipe.

The rate of ratcheting calculated using finite element analyses agrees more closely with data then the approximate solutions of Edmunds and Beer and Beaney. The approximations are conservative and over predict ratcheting strains. Beaney's equation was found to be the most accurate.

The ELBOW31 element available in the ABAQUS finite element library was used to predict the ratcheting in a recently completed test of a cyclically loaded pressurized elbow in Darmstadt, Germany. Correlation of finite element predictions with the test data was excellent when the strains were in the elastic range. When the applied load caused plastic strains, conservative estimates were calculated. Shortcomings in the kinematic hardening model which is based on a bilinear elastic-plastic material and uncertainties in the assumed yield stress and tangent modulus are believed to contribute to deviations between test and analysis.

References:

[1]   Severud, K. L., Anderson, M. J., Lindquist, M. R.,
       Wagner, S. E., Weiner, E. O., "High Level Seismic
       Response and Failure Prediction Methods for Piping,"
       Westinghouse Handford Co., Report No NUREG/CR-5023, WHC-
       EP-0081, RM, January 1988.

[2]   Lindquist, M. R., Anderson, M. J., Severud, L. K.,
       Weiner, E. O., "High-Level Dynamic Testing and
       Analytical Correlations for a Small Bore Piping System,"
       Jr. of Pressure Vessel Technology Vol. 110, August 1988,
       pp 270-75 and PVP-Vol 108, 1986 pp 169-83.

[3]   English, W. F., "Piping and Fitting Dynamic Reliability
       Program – First Semiannual Progress Report," May-
       October 1985, NEDC-31272, EPRI Contracy RP 1543-15,
       November 1985.

[4]   English, W. F., "Piping and Fitting Dynamic Reliability
       Program – Second Semiannual Progress Report," November-
       April 1985, NEDC-31364, EPRI Contracy RP 1543-15,
       November 1986.

[5]   Edmunds, H. G., Beers, F. J., "Notes on Incremental
       Collapse in Pressure Vessels," Journal of Mechanical
       Engineering Science, Vol. 3, No. 3, 1961, pp.187-199.

[6]   Beaney, E. M., "Failure of Pipework Subjected to Seismic
       Loading," Central Electricity Generating Board,
       Berkeley, Gloucestershire, UK, Report RD/BXXXX/R89.

[7]   Beaney, E. M., "Plasticity in Reonant Pipework
       Vibration," International Conference, "Pipework
       Engineering and Operation," Institution of Mechanical
       Engineers, London, UK, Feb. 21-22, 1989, pp 281-90.

[8]   Beaney, E. M., "Response of Pipes to Seismic
       Excitation – Effect of Pipe Diameter/Wall Thickness
       Ratio and Material Properties," Central Electricity
       Generating Board, Berkeley Nuclear Laboratories,
       Report TPRD/B/0637/N85, Berkeley, Gloucestershire, UK,
       GL 13 9PB.

[9]   Beaney, E. M., "Response of Pipes to Seismic
       Excitation – Effect of Pipe Diameter/Wall Thickness
       Ratio and Material Properties," Central Electricity
       Generating Board, Berkeley Nuclear Laboratories,
       Report TPRD/B/0637/N85, Berkeley, Gloucestershire, UK,
       GL 13 9PB.

[10]   D. R. Miller, "Thermal-Stress Ratchet Mechanism in
       Pressure Vessels," Transaction ASME, Jr. of Basic Engr.,
       June, 1959, pp 190-196.

[11] Gau, J. S., "Elastic-Plastic Behavior of Pressurized Pipe," Dissertation, Dept. of Mechanical Engineering, The Univ. of Akron, Akron, Ohio, Jan. 1990.

[12] Scavuzzo, R. J. and Lam, P. C., "Investigation of Design Criteria for Dynamic Loads on Nuclear Power Piping," Welding Research Bulletin 324, June, 1987.

[13] Scavuzzo, R. J., Lam, P. C. and Gau, J. S., "Finite Element Modeling of Elastic-Plastic Bending of Beams" 60th Shock and Vibration Bulletin, November 1989.

[14] Scavuzzo, R. J., Lam, P. C., Gau, J. S., "Experimental Studies of Ratcheting of Pressurized Pipe" Jr. of Pressure Vessel Technology, Vol. 113, May 1991 pp 210-218.

[15] Scavuzzo, R. J., Lam, P. C., Gau, J. S., " Ratcheting of Seismically Loaded Pressurized Pipe," ASME PVP - Vol. 182, July 1989, pp 103-10.

[16] Hibbitt, Karlsson and Sorensen, Inc.,"ABAQUS User Manual," Providence, RI 1988.

[17] Azzam, T. J., "Ratcheting in Cyclically Loaded Pressurized Elbows," Dissertation, The Unioversity of Akron, May 1992

[18] Azzam, T. J., Scavuzzo, R. J., " Elbow Ratcheting - A Comparison with Data," Proceedings of ICVPT 7, Volume 2, Dusseldorf, Germany, May 31 - June 5, 1992, pp 1445-54.

[19] Azzam, T. J., Scavuzzo, R. J., "Analysis of Ratcheting in Pressurized Pipe Elbows," PVP-Vol. 220, 1991, pp. 107-116.

[20] Jorg van Kann, Darmstadt, Germany (not published).

297
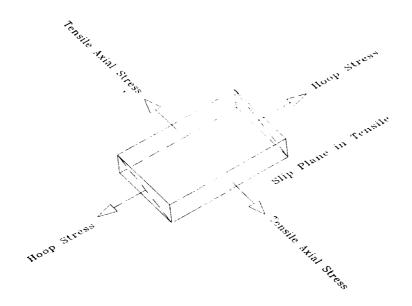
Fig. 1 Slip planes with axial tensile stresses
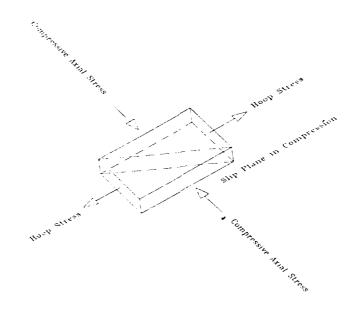
Fig. 2 Slip planes with axial compressive stresses
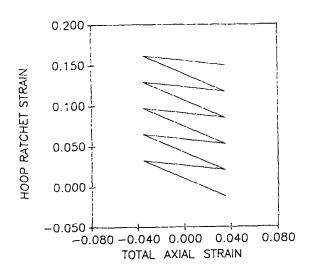
Fig. 3 Hoop ratcheting plotted against the cyclic axial
strain based on the Edmunds-Beer model [5].
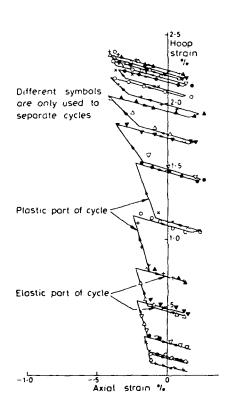
Fig. 4 Hoop ratcheting plotted against the cyclic axial
strain based on the experimental work of Beaney.

299
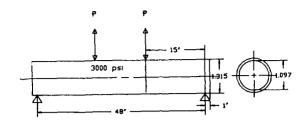
Fig. 5   Four point loading used in both experimental and analytical work.
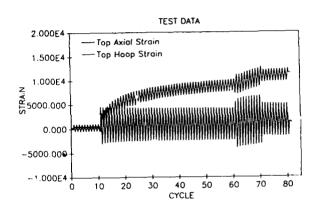
Fig. 6 Specimen 3,
Carbon Steel, Strain -μin./in.
Pressure = 3,000 psi,
Loading 1, Displacement = 0.25",
Center Deflection = 0.533
Loading 2, Displacement = 1.0",
Center Deflection = 2.676"
Loading 3, Displacement = 1.5",
Center Deflection = 3.911",
Maximum Load = 1,698 lbs.
Loading 4, Displacement = 1.0",
Center Deflection = 2.531"
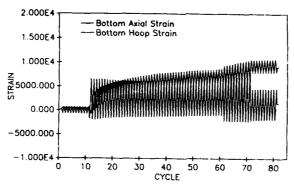
Fig. 7 Specimen 4,
304 SS, Strain -μin./in.
Pressure = 3,000 psi,
Loading 1, Displacement = 0.25",
Center Deflection = 0.533"
Loading 2, Displacement = 1.0",
Center Deflection = 2.570"
Loading 3, Displacement = 1.5",
Center Deflection = 3.911",
Maximum Load = 1,645 lbs.
Loading 4, Displacement = 1.0",
Center Deflection = 2.422"
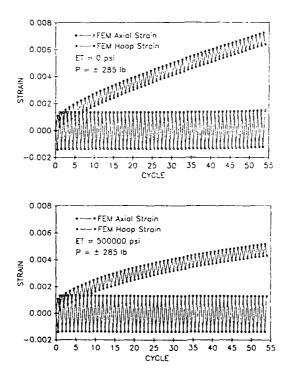
Fig. 8 Elastic-plastic finite element analysis of ratcheting
with an internal pressure of 3,000 psi and the
geometry of Figure 4.

301
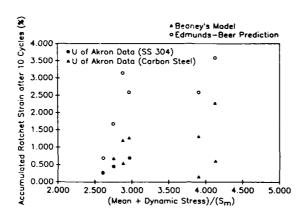
Fig. 9 Comparison of hoop ratcheting strain with predictions.

Fig. 10 Elbow test setup.

302

0 = Intrados
180 = Extrados
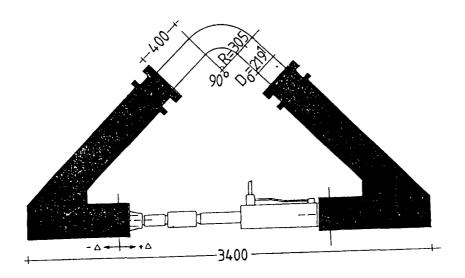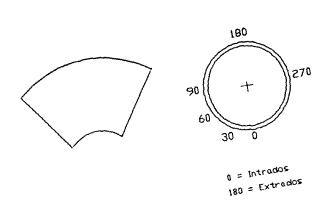
Axial strain measured at 0,30,60,90,180,270
Hoop strain measured at 90 and 270

Fig. 11 Strain gage.

APPLIED DISPLACEMENT
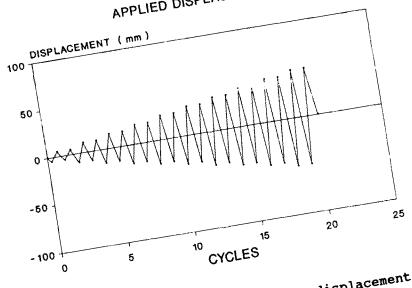
Fig. 12 Applied cyclic displacement.

303

Elements 1 - 10 and 28 - 37 are B31 beam elements
Elements 11 - 27 are ELBOW31 elbow elements

* NODE
○ ELEMENT
▭ Flange
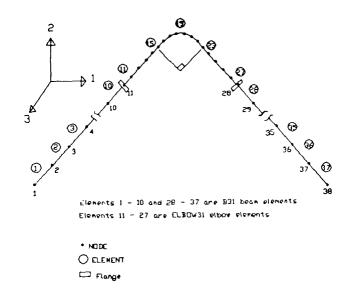
Fig. 13   Finite element model.

AXIAL STRAIN AT 0 DEGREES
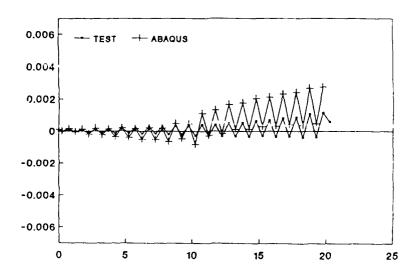
Fig. 14   Calculated and measured axial strains at 0°.

304

AXIAL STRAIN AT 30 DEGREES
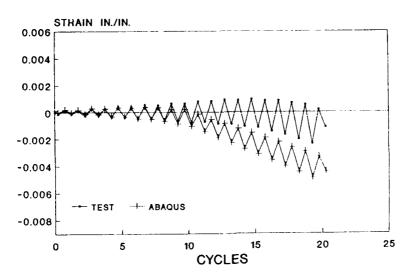
Fig. 15 Calculated and measured axial strains at 30°.

AXIAL STRAIN AT 60 DEGREES
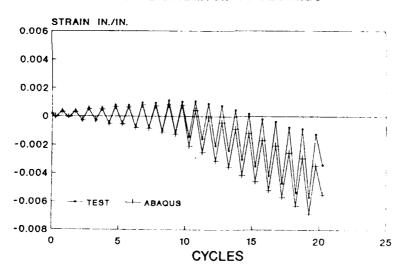
Fig. 16 Calculated and measured axial strains at 60°.

305

AXIAL STRAIN AT 90 DEGREES

STRAIN IN./IN.
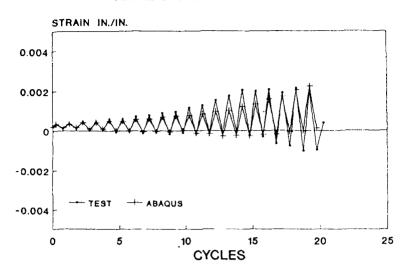
Fig. 17 Calculated and measured axial strains at 90°.

HOOP STRAIN AT 90 DEGREES

STRAIN IN./IN.
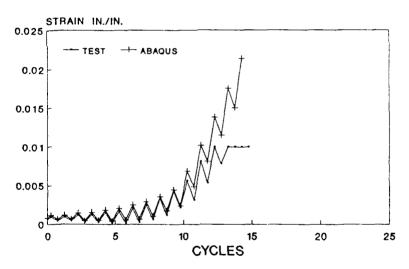
Fig. 18 Calculated and measured hoop strains at 90°.

306

HOOP STRAIN AT 270 DEGREES
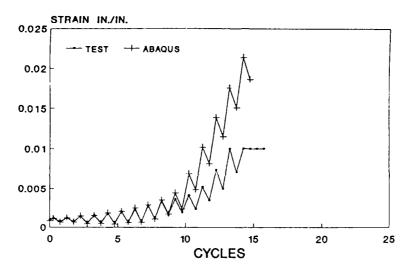
Fig. 19 Calculated and measured hoop strains at 270°.

307

METALLURGICAL EXAMINATION OF FAILED SUSPENSION LUGS

Marc Pepi
Victor K Champagne, Jr.
Gary Wechsler
Brendan McCahill
William Green
Phil Bennett
Walter Roy
U.S. Army Research Laboratory
405 Arsenal Street
AMSRL-MA-CB
Watertown, Massachusetts 02172

Abstract: Three Naval MS3314 suspension lugs failed during routine proof load testing. A failure analysis of two of the three broken lugs was performed. Visual examination of the lugs revealed a blackened region near the crack origin. The lugs were fabricated from steel conforming to the governing specification, as determined by chemical analysis. Metallographic analysis revealed the microstructure to consist of tempered martensite. Hardness tests revealed the parts conformed to the governing specification. Electron microscopy of failed lugs revealed that the black region exhibited features consistent with a high temperature oxide. Energy dispersive spectroscopy of the blackened region revealed a large oxygen concentration. It was concluded that the cause of failure was the result of forging laps formed during the fabrication of the lugs and the blackened region was the result of a tempering operation.

The in-process inspection procedures of the lug manufacturer were reviewed and found to be inadequate. Magnetic particle testing of lugs in inventory and in service revealed defects similar to those on the failed lugs. All of these defective lugs had been previously 100% magnetic particle tested and accepted by the manufacturer. Recommendations which improved these inspection procedures were presented to the manufacture of the lugs and incorporated into the in-process inspection procedures. In addition, all existing lugs were designated to be reinspected or replaced.

Key Words: Failure analysis; forging laps; high strength steel; magnetic particle inspection; nondestructive testing; suspension lugs.

Introduction: The Naval MS3314 suspension lug is utilized to secure the MK 82 series 1000-pound general purpose bomb to various Naval and Air Force fixed-wing aircraft. The lugs are fabricated from steel according to MIL-S-5000, and hardened to 38-44 HRC per MIL-H-6875. A closed die, hot forging process is utilized by the forging subcontractor in the manufacture of the lugs. The subcontractor subsequently performs a 100% magnetic particle inspection of the lugs. Threads are machined on

the components after the forging process and heat treatment by the primary contractor. The primary contractor then performs a 100% magnetic particle inspection on the finished lugs. The parts are subsequently cadmium plated according to QQ-P-416, Type II, Class 2. There are approximately 900,000 of these lugs currently in service or storage.

During routine proof load testing conducted at the manufacturing facility, three lugs failed before the mandatory time requirement at the designated tensile load was achieved, as specified in the applicable Automated Data List (ADL). Two of the three failed lugs were immediately sent to ARL to be examined for cause of failure. One of these lugs failed during the 6-degree angle proof test before the specified one minute hold time at 35,000 pounds tension was achieved. The other lug failed during the 35-degree angle proof load test 20 seconds into the specified one minute hold time at 24,000 pounds. ARL was requested by the MS3314 project managing office, the Naval Air Warfare Center (NAWC), to identify the cause of premature failure of the two lugs and determine if lugs in the field and inventory were also defective.

The following analytical tests and inspection procedures were performed:
- Failure analysis including visual examination/light optical microscopy, chemical analysis, metallographic analysis, micro- and macrohardness testing, electron microscopy and energy dispersive spectroscopy.
- In-process review of the magnetic particle inspection procedures of the lug manufacturers.
- Magnetic particle inspection of lugs in the field and in Naval inventory.

FAILURE ANALYSIS
Visual examination: Visual examination of the two failed lugs revealed a fracture through the entire cross-section of the lug handle also referred to as a bail. In one instance, the crack initiation site was located at the top of the bail where there is a large tensile stress during loading, as shown in Figure 1. The second failure occurred on the side of the bail, as shown in Figure 2. Both lugs contained a blackened fracture surface near the crack origin. The black color of these regions suggested that they may have been exposed to elevated temperatures, most likely during the tempering operation. Therefore, the blackened areas most likely represented a heat treatment scale. When both fracture halves of each failure were placed together, the pattern of scale matched and a surface lap was observed. Further examination of the as-forged surface of the first lug revealed another lap on the top of the bail. The defect was approximately 0.10 inches long and situated parallel to the fracture plane. The lap was believed to be the result of the forging process. No laps or other significant surface discontinuities were noted on the external surface of the second failed lug.

Chemical Analysis: Atomic absorption and inductively coupled argon plasma emission spectroscopy were utilized to determine the chemical composition of material sectioned from the failed lugs. The carbon and sulfur content was analyzed by the Leco
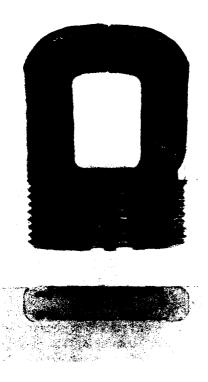
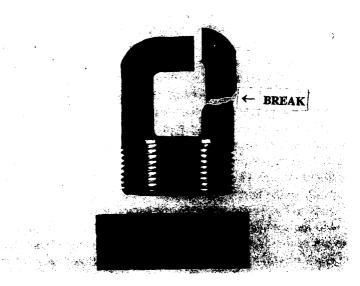**Figure 1**   The first failed suspension lug shown in the as-received condition. 1X

Figure 2   The second failed suspension lug shown in the as-received condition.
Reduced 30%

311

combustion method. The compositional ranges of the material representing the two failed lugs compared favorably with the governing specification.

Metallographic Examination: Metallographic samples were taken through the cross-section of the fracture origin on both failures. The general cleanliness of the material was relatively good, despite some evidence of scattered inclusions and manganese sulfide stringers. The microstructure adjacent to the laps etched slightly darker which may be indicative of light carburization. Investigation of as-forged surfaces revealed a similar etching characteristic. Carburization would only occur to those surfaces exposed directly to elevated temperatures during heat treatment. Therefore, the surfaces of the laps were most likely exposed to the heat treat atmosphere. The microstructure of both lugs consisted of fine tempered martensite, as shown in Figure 3. This general structure was observed throughout the thickness of the material, and is representative of an austenitized, quenched and tempered steel. Flow lines, indicative of prior forming operations, were also observed .

Hardness Testing: A series of micro- and macrohardness measurements were performed on cross-sections of the failed lugs. The Knoop microhardness, and the Rockwell "C" macrohardness scales were utilized. The required hardness as specified on the governing engineering drawing of the component was HRC 38-43. The hardness values obtained conformed to the specification.

Electron Microscopy/Energy Dispersive Spectroscopy: The darkened surfaces previously observed at the crack origins of both failed lugs were analyzed by energy dispersive spectroscopy (EDS). Figure 4 illustrates a representative spectra obtained from these regions. The large iron peak and oxygen peak indicate a corrosion product or heat treat scale. The fracture surfaces located away from the crack initiation sites yielded EDS spectrum with no significant concentration of oxygen, shown in Figure 5. The mode of fracture of each failed lug was analyzed through a scanning electron microscope (SEM). Figure 6 contains a full view of a fracture half from the first failure, which was similar in nature to that of the second failure. Oblique lighting was utilized to accentuate the fractographic features of this surface. The surface was divided into three distinct zones. Zone 1 represents the darkened region, which was determined to have been the result of a forming defect. Zone 2, which encompasses most of the entire fracture surface, was caused by overload, as determined by the presence of ductile dimples. This was the anticipated mode of failure, due to the circumstance leading to the failures (proof load testing). Zone 3 also fractured by overload conditions and represents the shear lip region where cracking occurred at a 45-degree angle to the applied stress. The radial lines and chevron pattern converge at a region on the top of the suspension lug handle adjacent to the blackened surface identifying the crack origin, as denoted by the arrow. A closer examination of the darkened surfaces was performed, which revealed a featureless condition associated with oxide formation.
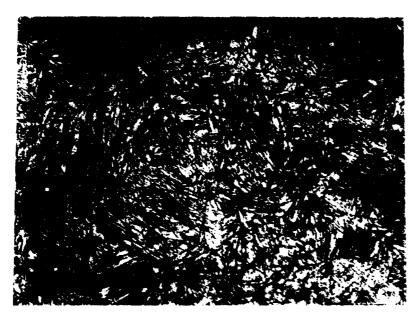
Figure 3   The microstructure of the failed lugs consisted of a fine tempered
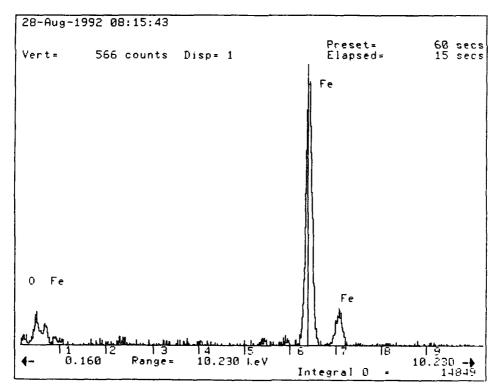martensite. 1000X

Figure 4   Representative EDS spectrum obtained from the blackened regions at the
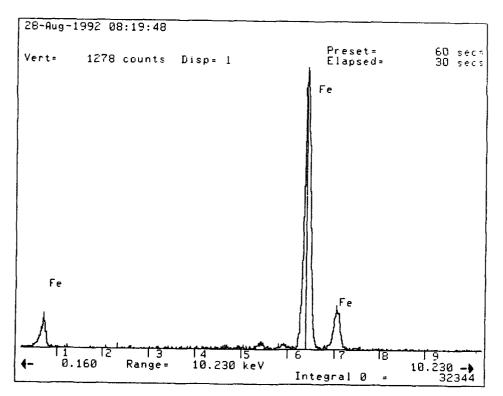crack initiation sites.

```
28-Aug-1992 08:19:48

Vert=    1278 counts  Disp= 1          Preset=      60 sec
                                       Elapsed=     30 sec

                                            Fe




  Fe
                                              Fe


      1    2    3    4    5    6    7    8    9
  ←   0.160   Range=   10.230 keV                  10.230 →
                                 Integral 0  =       32344
```

Figure 5   Representative EDS spectrum obtained from an area of the fracture surface away from the crack initiation sites.
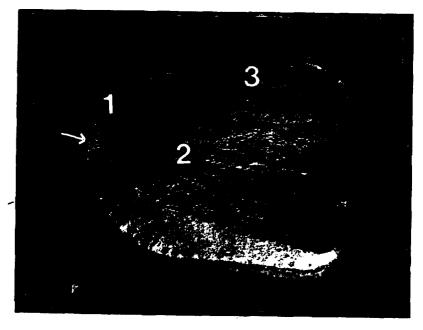
Figure 6   Macrograph of representative fracture half of a failed lug. 7.5X

314

IN-PROCESS REVIEW OF THE MANUFACTURER INSPECTION PROCEDURE

Results from the failure analysis suggested that the presence of laps was the primary cause of premature failure of the two lugs during proof load testing. Both the contractor and subcontractor of the lugs utilize magnetic particle inspections, as required by the ADL, to identify and reject lugs with detrimental defects such as laps. To determine if the requirements for 100% magnetic particle inspection were satisfied, a review of the magnetic particle process was performed. In addition, a sample of in service lugs from various heat treatment lots were re-inspected to ascertain the percentage of defective lugs currently in-service.

Magnetic Particle Process Review: The magnetic particle process review consisted of an analysis of the written standard operation procedures used by contractors and an on-site visual inspection of the magnetic particle inspection facilities.

The minimum requirements for magnetic particle examination of the MS3314 lug as stated in section 4.4.6 of the MS3314 ADL are as follows:

4.4.6 MAGNETIC PARTICLE TEST BEFORE PLATING - SUSPENSION LUGS SHALL BE TESTED AS SPECIFIED IN MIL-STD-1949A EXCEPT THAT THE LUGS SHALL BE MAGNETIZED IN TWO PLANES. IN THE FIRST PLANE, THE LUGS SHALL BE HUNG ON A BAR THROUGH THE LOOP AND THE BAR ENDS CLAMPED BETWEEN THE ELECTRODES. THE LUGS SHALL THEN BE MAGNETIZED WITH 1000 TO 1200 AMPERES. UPON COMPLETION OF THE FIRST MAGNETIZATION, THE LUGS SHALL THEN BE MAGNETIZED THROUGH THE BASE PERPENDICULAR TO THE THREADS. THE LUGS SHALL BE INSPECTED AS SPECIFIED IN MIL-STD-1949 AFTER THEY HAVE BEEN MAGNETIZED IN EACH PLANE. THE WET FLUORESCENT PROCEDURE AND CONTINUOUS APPLICATION METHOD SHALL BE USED.

The review of the magnetic particle inspection process of the primary contractor identified the following discrepancies:
1) The written procedure did not clearly define the sequence of operations, (i.e. direction of magnetism, order of magnetism), as required by MIL-STD-1949A.
2) The inspection technique incorporated the use of Duovec technology which, in theory, can magnetize each lug in all directions with one shot thus eliminating the need for a second shot. This technology had been used previously in private industry and was accepted by government personnel for use on the suspension lugs. However, analysis by ARL personnel found the Duovec equipment used by the contractor incapable of locating defects oriented parallel to the lug parting line. Furthermore, reducing the inspection procedure from two shots to one shot reduced the time each lug was visually inspected.

315

The review of the magnetic particle inspection process of the forging subcontractor identified the following discrepancies:

1) The written procedure described a method for magnetizing the lugs, referred to as the coil shot, which did not meet the requirements for coil shot magnetization as specified in MIL-STD-1949A.

2) The ADL required that all lugs shall be magnetized through the base perpendicular to the threads. The written procedure described a method of magnetization, referred to as a head shot, which met this requirement. However, the written procedures further stated that this method of magnetization was optional and that the improper coil shot described above was the primary method of magnetization.

3) Visual inspection of the magnetic particle technique identified a relatively long time delay between magnetization of the lugs and actual visual inspection of the lugs. This time delay could result in loss of test sensitivity and poor test results. Poor lug handling technique was also observed at the facility of the subcontractor which could further reduce the overall sensitivity of the inspection process.

Magnetic Particle Sample Inspection: To determine the percentage of defective lugs currently in service, a random sample of 4,050 service lugs were magnetic particle inspected by ARL. The lugs were divided into eighteen (18) separate heat treatment lots and were selected in accordance with MIL-STD-105: "Sampling Procedures and Tables for Inspection by Attributes". The magnetic particle inspection was performed in accordance to a procedure developed by ARL which meets all requirements of the MS3314 ADL and MIL-STD-1949A.

A total of 105 defective lugs were detected which computes to a reject rate of 2.6%. Defects found on the lugs included 45 seams, 28 laps, 9 gouges and 23 small bright indications located on the machined surfaces of the lug handles. The seams were located primarily on the inside and outside surface on the two vertical lug handles and were oriented parallel to the parting line (Figure 7). The laps were located on the top of the bail and the inside corners of the bail (Figure 8). The gouges were located in the same areas as the laps. Defect sizes ranged from 1-7mm for the seams, 2-5mm for the laps, 3-5mm for the gouges and under 1mm for the small bright indications.

CONCLUSIONS
Failure Analysis:

1) Visual examination revealed a blackened region at the crack origin. In addition, a forming lap was found on the external bail surface of one of the failed lugs.

2) Material sectioned from the failed lugs and subjected to chemical analysis conformed to the governing specification.

3) Metallographic examination adjacent to the blackened regions at the crack initiation sites showed slight carburization upon etching. This indicated that the blackened regions were exposed to the high temperatures associated with the heat treatment. The microstructure was generally clean, and consisted of a fine tempered martensite.

4) Hardness testing performed on samples sectioned from the failed lugs revealed results within the governing specifications.
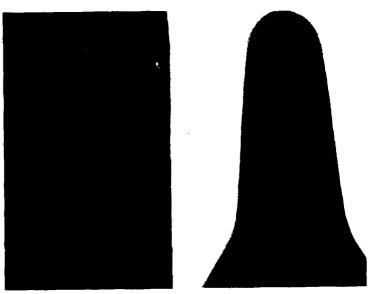
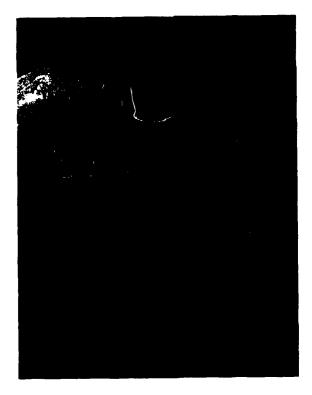**Figure 7**   Black light macrograph of a typical seam defect noted on lugs in inventory. 3X

Figure 8   Blacklight macrograph of typical lap defect noted on lugs in inventory. 4X

317

5) Electron microscopy of the blackened surface revealed a featureless condition associated with oxide formation. It was concluded that the lugs failed due to overload conditions, as determined by the predominantly ductile dimpled fracture surface. Energy dispersive spectroscopy of the blackened regions revealed evidence of a corrosion product or a heat treat scale.

6) In summary, the dark region located at the fracture origins were exposed to elevated temperatures, most likely during the tempering operation. The oxide present on the internal surface of the defect was evidence supporting this claim. It is believed that forging laps existed prior to the heat treatment and that crack initiation occurred at these defects during proof load testing. The surface laps may have caused the lugs to fail at lower loads due to the affects of stress concentration at the root of these defects.

In-Process Review Of The Manufacturer Inspection Procedure:
1) The written procedure used by the primary contractor did not meet the requirements as specified by the governing military document, MIL-STD-1949A.

2) The Duovec inspection system used by the primary contractor was not capable of detecting seam discontinuities. This system also reduced visual inspection time for each lug thereby increasing the possibility of missed defects.

3) The inspection procedure used by the forging subcontractor was not authorized by the ADL and did not meet the requirements of MIL-STD-1949A. Poor lug handling practiced by the subcontractor further reduced the sensitivity of the inspection.

4) A total of 105 defective lugs were detected during the sample inspection of the 4050 lugs in inventory. This computed to a reject rate for in service lugs of 2.6%.

Recommendations: As discussed previously in the magnetic particle process review section, ARL has developed a magnetic particle procedure which meets all the requirements of the ADL and MIL-STD-1949A. This procedure has been utilized successfully by ARL during sample inspection in detecting all detrimental surface defects. This procedure has since been incorporated into the inspection process of both of the contractors. In addition, an inspection production line utilizing the ARL procedure will be established at an Army depot to reinspect all lugs currently in service or in storage.

A GENERALIZED METHOD FOR SIMULATING INDUCTION MACHINE DYNAMIC CURRENT CHARACTERISTICS INCLUDING NONIDEAL EFFECTS

Randy R. Schoen and Thomas G. Habetler
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, GA 30332-0250

Abstract: Conventional stationary reference-frame theory is used to transform the voltage equations of an ideal 2-pole, 3-phase induction machine into an equivalent orthogonal 2-phase set of time-varying differential equations. These are easily solved using numerical computation software packages. Since the dynamic model of the motor includes the mechanical equation, any arbitrary time function of load torque can be specified from which the resulting stator current is calculated. The dynamic model can also be used to determine the stator current in the presence of a machine fault. From the standpoint of a theoretical analysis, most machine faults can be grouped into two categories: 1) those which result in a torque or speed oscillation of the machine, and 2) those which cause an anomaly in the air gap flux distribution. This paper will show an easy and simple approach to determining the instantaneous stator current for an idealized machine operating under these conditions and/or with any arbitrary load.

Key Words: Computer Simulation, Dynamic Modeling, Induction Motors, Motor Current Analysis.

Introduction: The goal of this work is to model and simulate an induction machine operating under any arbitrary time-varying load conditions in the presence of a non-uniform magnetic field. Various types of fault conditions in induction motors cause the magnetic field in the air gap of the machine to be nonuniform. The effects of these faults can be simulated by modeling the fault as a harmonic component in the stator magnetic field. Other types of motor and load faults (such as a worn bearings) cause the load torque of the machine to be modulated. Both of these types of fault categories can be simulated using the dynamic, time-varying equations of the induction machines.

In order to analyze the harmonic content in the current and flux of an induction machine, the time-varying differential equations which describe the magnetically coupled, three-phase, stator and rotor windings must be analyzed. The work of Stanley [1], Kron, Krause and Thomas[2] applied the Reference Frame Theory to induction machines in order to greatly reduce the complexity of the equations which describe the machine. This is accomplished by transforming the three-phase windings to an equivalent set of orthogonal two-phase windings.

This paper begins with the development of the 3-phase voltage equations for an ideal 2-pole induction motor. Stationary Reference Frame Theory is then used to transform these equations into an equivalent orthogonal 2-phase set of time-varying differential equations. These equations, combined with the electromechanical torque equation and the mechanical dynamic equation, are used to simulate machine operation with both a constant and an arbitrary time-varying load torque.

Determination of the machine inductances is then examined and used to model an anomaly in the machine's air gap flux distribution. In particular, the inductances for a rotating air

319

gap eccentricity are determined and used to simulate its effect upon the machine operation. A second simulation, including both an air gap eccentricity and a sinusoidally varying load torque, is also presented.

Dynamic Model of Induction Machine: The idealized two-pole, three phase induction motor used to develop the dynamic equations is graphically represented in Figure 1(a). It is based on the following assumptions:

1) smooth rotor and uniform air gap
2) negligible magnetic saturation and core losses
3) sinusoidally distributed windings producing a sinusoidal flux distribution
4) squirrel-cage rotors can be represented by an equivalent 3-phase winding
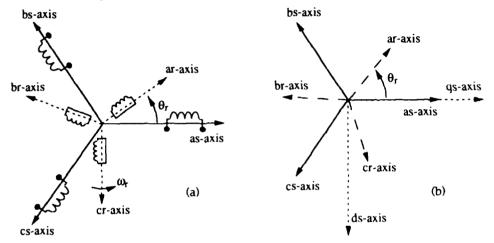
Figure 1. (a) Three phase equivalent windings, (b) abc- and dq-coordinate axis.

Using this model, the voltage equations for each of the stator and rotor windings can be written as:

$$\begin{bmatrix} \mathbf{V}_{abcs} \\ \mathbf{V}_{abcr} \end{bmatrix} = \begin{bmatrix} r_s & 0 \\ 0 & r_r \end{bmatrix} \begin{bmatrix} i_{abcs} \\ i_{abcr} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \lambda_{abcs} \\ \lambda_{abcr} \end{bmatrix}$$

(1)

where the instantaneous voltage, current, and flux linkage values are:

$$\mathbf{V}_{abcs} = \begin{bmatrix} V_{as} \\ V_{bs} \\ V_{cs} \end{bmatrix}, \quad i_{abcs} = \begin{bmatrix} i_{as} \\ i_{bs} \\ i_{cs} \end{bmatrix}, \quad \lambda_{abcs} = \begin{bmatrix} l_{as} \\ l_{bs} \\ l_{cs} \end{bmatrix}$$

(2)

$$\mathbf{V}_{abcr} = \begin{bmatrix} V_{ar} \\ V_{br} \\ V_{cr} \end{bmatrix}, \quad i_{abcr} = \begin{bmatrix} i_{ar} \\ i_{br} \\ i_{cr} \end{bmatrix}, \quad \lambda_{abcr} = \begin{bmatrix} l_{ar} \\ l_{br} \\ l_{cr} \end{bmatrix}$$

(3)

and the stator and rotor resistances are expressed, in terms of the stator and rotor phase resistances (r_s, r_r), by:

320

$$\mathbf{r}_s = \begin{bmatrix} r_s & 0 & 0 \\ 0 & r_s & 0 \\ 0 & 0 & r_s \end{bmatrix}, \mathbf{r}_r = \begin{bmatrix} r_r & 0 & 0 \\ 0 & r_r & 0 \\ 0 & 0 & r_r \end{bmatrix} \tag{4}$$

The flux linkages are defined to be :

$$\begin{bmatrix} \lambda_{abcs} \\ \lambda_{abcr} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_s & \mathbf{L}_{sr} \\ (\mathbf{L}_{sr})^T & \mathbf{L}_r \end{bmatrix} \begin{bmatrix} \mathbf{i}_{abcs} \\ \mathbf{i}_{abcr} \end{bmatrix} \tag{5}$$

where the stator and rotor inductances are given by:

$$\mathbf{L}_s = \begin{bmatrix} L_{ls}+L_{ms} & -\frac{1}{2}L_{ms} & -\frac{1}{2}L_{ms} \\ -\frac{1}{2}L_{ms} & L_{ls}+L_{ms} & -\frac{1}{2}L_{ms} \\ -\frac{1}{2}L_{ms} & -\frac{1}{2}L_{ms} & L_{ls}+L_{ms} \end{bmatrix} \tag{6}$$

$$\mathbf{L}_r = \begin{bmatrix} L_{lr}+L_{mr} & -\frac{1}{2}L_{mr} & -\frac{1}{2}L_{mr} \\ -\frac{1}{2}L_{mr} & L_{lr}+L_{mr} & -\frac{1}{2}L_{mr} \\ -\frac{1}{2}L_{mr} & -\frac{1}{2}L_{mr} & L_{lr}+L_{mr} \end{bmatrix} \tag{7}$$

$$\mathbf{L}_{sr} = L_{sr} \begin{bmatrix} \cos(\theta_r) & \cos\left(\theta_r + \frac{2\pi}{3}\right) & \cos\left(\theta_r - \frac{2\pi}{3}\right) \\ \cos\left(\theta_r - \frac{2\pi}{3}\right) & \cos(\theta_r) & \cos\left(\theta_r + \frac{2\pi}{3}\right) \\ \cos\left(\theta_r + \frac{2\pi}{3}\right) & \cos\left(\theta_r - \frac{2\pi}{3}\right) & \cos(\theta_r) \end{bmatrix} \tag{8}$$

These inductances are given in terms of the stator leakage and mutual inductances (L_{ls}, L_{ms}), the rotor leakage and mutual inductances (L_{lr}, L_{mr}), and the stator-rotor mutual inductance (L_{sr}).

These differential equations are difficult to solve because of the numerous coupled magnetic circuits. Stationary Reference Frame Theory [3,4] reduces the complexity of these equations by mapping the stator and rotor 3-phase axis onto a set of three orthogonal axis, labeled direct (ds,dr), quadrature (qs,qr) and zero-sequence (0s,0r). The relationship between the coordinate systems is shown in Figure 1(b), where the zero-sequence axis points out of the page.

The transformations between the stator abc-variables and their equivalent qd0-variables are:

$$[\chi_{qd0s}] = [\mathbf{T}_s][\chi_{abcs}] \quad \text{and} \quad [\chi_{abcs}] = [\mathbf{T}_s]^{-1}[\chi_{qd0s}] \tag{9}$$

where

$$[T_s] = \frac{2}{3}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad [T_s]^{-1} = \begin{bmatrix} 1 & 0 & 1 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 1 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & 1 \end{bmatrix} \tag{10}$$

Likewise, the transformations between the rotor abc-variables a.id their equivalent qd0-variables are:

$$[\chi_{qd0s}] = [T_r][\chi_{abcs}] \quad \text{and} \quad [\chi_{abcs}] = [T_r]^{-1}[\chi_{qd0s}] \tag{11}$$

where

$$[T_r] = \frac{2}{3}\begin{bmatrix} \cos(\theta_r) & \cos\left(\theta_r + \frac{2\pi}{3}\right) & \cos\left(\theta_r - \frac{2\pi}{3}\right) \\ -\sin(\theta_r) & -\sin\left(\theta_r + \frac{2\pi}{3}\right) & -\sin\left(\theta_r - \frac{2\pi}{3}\right) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{12}$$

$$[T_r]^{-1} = \begin{bmatrix} \cos(\theta_r) & -\sin(\theta_r) & 1 \\ \cos\left(\theta_r + \frac{2\pi}{3}\right) & -\sin\left(\theta_r + \frac{2\pi}{3}\right) & 1 \\ \cos\left(\theta_r - \frac{2\pi}{3}\right) & -\sin\left(\theta_r - \frac{2\pi}{3}\right) & 1 \end{bmatrix} \tag{13}$$

Applying these transformation matrices to the machine voltage and flux linkage equations, (1) and (5), yields:

$$\begin{aligned}
\frac{d}{dt}\lambda_{qs} &= V_{qs} - r_s i_{qs} & \lambda_{qs} &= L_{ls}i_{qs} + L_{ms}\left[\frac{3}{2}i_{qs} + \frac{3}{2}i_{qr}\right] \\
\frac{d}{dt}\lambda_{ds} &= V_{ds} - r_s i_{ds} & \lambda_{ds} &= L_{ls}i_{ds} + L_{ms}\left[\frac{3}{2}i_{ds} + \frac{3}{2}i_{dr}\right] \\
\frac{d}{dt}\lambda_{qr} &= V_{qr} - r_r i_{qr} + \omega_r\lambda_{dr} & \lambda_{qr} &= L_{lr}i_{qr} + L_{ms}\left[\frac{3}{2}i_{qs} + \frac{3}{2}i_{qr}\right] \\
\frac{d}{dt}\lambda_{dr} &= V_{dr} - r_r i_{dr} - \omega_r\lambda_{qr} & \lambda_{dr} &= L_{lr}i_{dr} + L_{ms}\left[\frac{3}{2}i_{ds} + \frac{3}{2}i_{dr}\right]
\end{aligned} \tag{14}$$

and

$$\begin{aligned}
\frac{d}{dt}\lambda_{0s} &= V_{0s} - r_s i_{0s} & \lambda_{0s} &= L_{ls}i_{0s} \\
\frac{d}{dt}\lambda_{0r} &= V_{0r} - r_s i_{0r} & \lambda_{0r} &= L_{lr}i_{0r}
\end{aligned} \tag{15}$$

It is easily seen that the number of coupled magnetic circuits has been significantly decreased. With the motor connected to a three-phase supply with no neutral wire, there are no zero-sequence voltages or currents in the machine, allowing the voltage and flux linkage equations to be reduced to the set defined by (14).

The electromagnetic torque developed by the machine may now be expressed in terms of the qd0-variables. At its most fundamental level, the electromagnetic torque is produced by the interaction of the total flux linking the stator windings and the MMF produced by the current flowing in the windings and may then be defined as the cross-product of these two

322

quantities. Because the actual stator quantities have been transformed to an equivalent set of orthogonal variables, the electromagnetic torque equation can be written as:

$$T_e = \frac{3}{2}\left(\lambda_{ds}i_{qs} - \lambda_{qs}i_{ds}\right)$$

(16)

Finally, the equation describing the mechanical dynamics of the machine and its load, neglecting friction and including both the machine and load inertia in J, is given by:

$$\frac{d}{dt}\omega_r = \frac{1}{J}\left(T_e - T_{load}\right)$$

(17)

These equations, (14), (16) and (17), are easily implemented using any numeric computation software with a differential equation solver. They also have the advantage that the A-phase stator current (i_{as}) is equal to the quadrature stator current (i_{qs}) and does not require an inverse transformation to be evaluated.

Simulation of Time-Varying Load Torque: Any load torque that can be expressed mathematically can be utilized in the simulation. These can include load torques of any shape that are dependent upon time or rotor position. To illustrate the effects of a time-varying load torque, the torque was modeled as a constant load with a 10% sinusoidal variation:

$$T_{load} = T_{avg}\left(1 + 0.1\ \cos\theta_r\right)$$

(18)

For comparison, simulations were performed for both a constant load and a sinusoidally varying load. Current spectrums were generated for both the load conditions and displayed in Figures 2(a) and 2(b). From equation (16), it can be seen that the stator current will

(a) Normalized current vs. Hertz

(b) Normalized current vs. Hertz
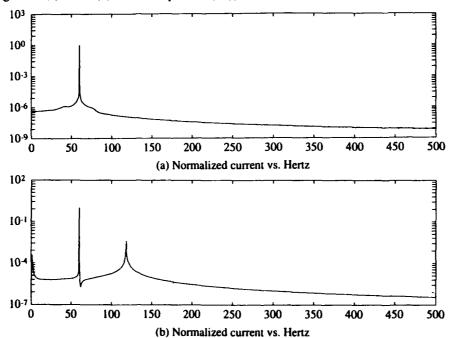
Figure 2. Normalized current spectrum for (a) constant load and
(b) sinusoidally varying load.

323

have frequency components at $60 \pm \omega_r$ Hz, since the stator flux has only a 60 Hz component.

Other parameters that may be simulated include the rotor position, the mechanical speed, the electromagnetic torque, shown in Figure 3(a), and the flux linkages, or their spectrums, shown in Figure 3(b).
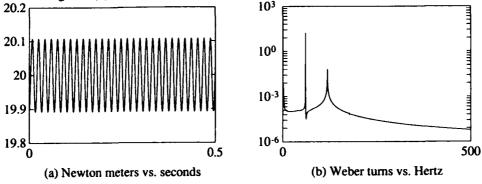

(a) Newton meters vs. seconds   (b) Weber turns vs. Hertz

Figure 3.(a) Electromagnetic torque and (b) rotor flux linkage spectrum
for a sinusoidally varying load.

Model of Air Gap Anomaly: Because the flux density in the air gap is defined as the product of the winding MMF and the air gap permeance, variations in either of these will generate anomalies in the flux distribution. For the ideal machine, the air gap flux density is perfectly sinusoidal because of the assumptions. However, this is not normally the case with harmonics caused by both the MMF and the permeance. The harmonics associated with the winding MMF are mainly determined by the winding distribution, however, the air gap permeance is dependent upon numerous effects including out-of-round rotors, unbalance, misalignment, and mechanical shaft vibrations caused by bearing or load faults. Regardless of the source, these anomalies have the same affect upon the flux density, and thus the machine inductances, and need only be considered once.

In order to understand how these variations affect the flux density, the steps required to calculate the inductances will be reviewed, using a sinusoidal winding distribution and a rotating air gap eccentricity as an example. The stator a-phase winding distribution, shown in Figure 4a, can be defined as:

$$N_{as} = \frac{N_s}{2} | \sin (\varphi_s) |$$

(19)

where φ_s is the angular measure around the air gap.

Because the current flow is defined to be out of the page for $0 \le \varphi_s \le \pi$ and into the page otherwise, the MMF produced by an instantaneous current, i_{as}, flowing through the winding is shown in Figure 4b and can be written as :

$$MMF_{as} = i_{as} \frac{N_s}{2} \cos (\varphi_s)$$

(20)

The other stator winding distributions are changed only by a shift in phase. while the rotor winding distributions must also include the change in rotor position, θ_r, in their expressions.

324
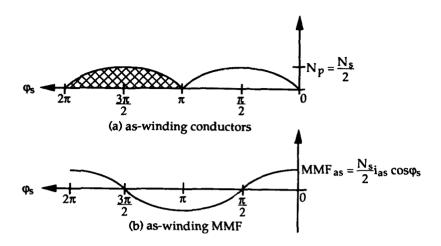
(a) as-winding conductors

(b) as-winding MMF

Figure 4. Stator phase (a) winding distribution and (b) produced MMF.

The flux density in the air gap due to current flowing in a winding is defined to be the product of the MMF and the permeance of the air gap. For the as-winding, this gives,

$$B_{as}(\varphi_s, \theta_r) = MMF_{as}(\varphi_s) \times P_{ag}(\varphi_s, \theta_r) \tag{21}$$

Permeance may be considered to be a conductance to the MMF which is produced by current flow in the winding and is proportional to the inverse of the length of the air gap. Under the initial assumptions, this permeance was constant because of the uniform air gap, however, any variation in the air gap can be easily modeled as a variation of the permeance. These variations can be expressed as a Fourier series and may be either stationary (22) or rotating (23).

$$P_{ag}(\varphi_s, \theta_r) = \delta_0 + \sum_n \delta_n \cos[n\, \varphi_s + \alpha_n] \tag{22}$$

$$P_{ag}(\varphi_s, \theta_r) = \delta_0 + \sum_n \delta_n \cos[n\,(\varphi_s - \theta_r) + \alpha_n] \tag{23}$$

A stationary eccentricity, like the one shown in Figure 5a, maintains the same position throughout time and can be described by,

$$P_{ag}(\varphi_s) = \delta_0 + \delta_1 \cos \varphi_s \tag{24}$$

The permeance of a rotating eccentricity changes over time since the rotor position moves from its initial position, Figure 5a, to a new position ($\theta_r = \omega t$), Figure 5b, at some later time, t. This variation does not need to be at rotational speed ($\omega = \omega_r$), but may be at any desired frequency. The equation that describes this variation at rotational speed, which will be used in the eccentric air gap simulation, can be expressed as,

$$P_{ag}(\varphi_s, \theta_r) = \delta_0 + \delta_1 \cos(\varphi_s - \theta_r) \tag{25}$$
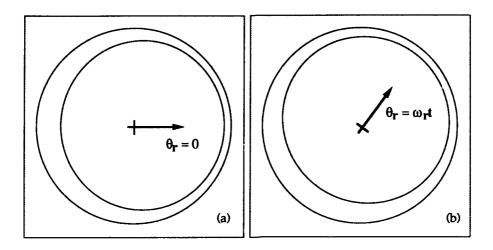
325

Figure 5. Depiction of air gap eccentricity.

Once the flux density (21) has been determined, it is possible to calculate the flux linking a single coil in the machine integrating over the surface of the coil. This is given by,

$$\Phi_{1sas}(\varphi_s,\theta_r) = \int_{\varphi_s}^{\varphi_s+\pi} B_{as}(\xi,\theta_r) \cdot r l \cdot d\xi$$

(26)

The flux linkage for an entire winding is then determined by summing the effects of each coil in the winding. For the self-inductance, there is an additional term, $L_{ls}i_{as}$, to account for the leakage inductance. The stator-stator flux linkages are given by,

$$\lambda_{asas} = L_{ls}i_{as} + \int_{\pi}^{2\pi} N_{as}(\varphi_s) \cdot \Phi_{1sas}(\varphi_s,\theta_r) \cdot r l \cdot d\varphi_s$$

(27)

where r is the mean radius of the air gap and l is the axial length of the rotor.

Using the definition of flux linkage (5), the as-phase winding self-inductance can be expressed as:
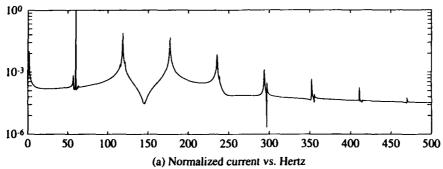
$$L_{asas} = \frac{\lambda_{asas}}{i_{as}} = L_{ls} + L_{ms}$$

(28)

These calculations were completed for all self- and mutual-inductances of a 3-phase machine with a rotating eccentricity (25) and the results transformed to the qd0-reference frame. The voltage equations (14) were unaffected, but the flux linkage equations were modified as follows:

326

$$\lambda_{qs} = L_{ls}i_{qs} + L_{ms}\left[\left(\frac{3}{2} - \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{qs} - \left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{ds} + \left(\frac{3}{2} - \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{qr} + \left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{dr}\right]$$

$$\lambda_{ds} = L_{ls}i_{ds} + L_{ms}\left[-\left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{qs} + \left(\frac{3}{2} + \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{ds} + \left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{qr} + \left(\frac{3}{2} + \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{dr}\right]$$

$$\lambda_{qr} = L_{lr}i_{qr} + L_{ms}\left[\left(\frac{3}{2} - \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{qs} + \left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{ds} + \left(\frac{3}{2} - \frac{3}{4}\delta\cdot\cos2\theta_r\right)i_{qr} - \left(\frac{3}{4}\delta\cdot\sin2\theta_r\right)i_{dr}\right]$$

$$\lambda_{dr} = L_{lr}i_{dr} + L_{ms}\left[\left(\frac{3}{4}\delta\cdot\sin\theta_r\right)i_{qs} + \left(\frac{3}{2} + \frac{3}{4}\delta\cdot\cos\theta_r\right)i_{ds} - \left(\frac{3}{4}\delta\cdot\sin2\theta_r\right)i_{qr} + \left(\frac{3}{2} + \frac{3}{4}\delta\cdot\cos2\theta_r\right)i_{dr}\right]$$

where $\delta = \delta_0 / \delta_1$. (29)

Simulation of Eccentric Air-gap: The modified flux linkage equations (29) were incorporated into the machine equations and used to simulate a sinusoidal air gap variation of 1% ($\delta = 0.01$). From equation (29), it can be seen that, because of the interaction of the stator and rotor inductances, the flux linkages will contain frequency components at multiples of the rotational speed. When these are included in the electromagnetic torque equation (16), it is apparent that, while under constant load torque, the stator current will contain multiples of the rotational speed. When these are included in the electromagnetic torque equation (16), it is apparent that, while under constant load torque, the stator current will have frequency components at $60 \pm n\omega_r$ Hz. These components are easily seen in the frequency spectrum of the phase current, Figure 6(a).

(a) Normalized current vs. Hertz
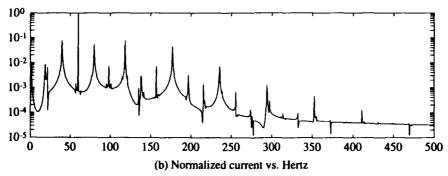

(b) Normalized current vs. Hertz
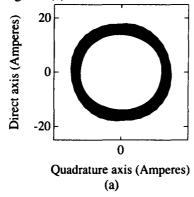
Figure 6. Normalized current spectrum for eccentric air gap with
(a) constant load and (b) sinusoidally varying load.

A second simulation of the air gap eccentricity was performed with a 20 Hertz sinusoidally varying load. The results of this simulation is shown in Figure 6(b). It can be seen from the figure that not only does the current have frequency components at 60 ± 20 Hz, but the

interaction of the torque oscillation and the eccentric air gap produces harmonics at $n\omega_r \pm$ 20 Hz.

Another representation of the variation in magnitude of the phase currents is shown in Figure 7. Here the direct-axis stator current is plotted against the quadrature-axis stator current. In the ideal machine with a constant load torque, this plot would generate a circle, however, because of the eccentricity the trajectory of the circle has increased. This is illustrated in Figure 7(a). When the sinusoidal load is added, the variation in the current magnitude increases and causing the trajectory of the circle to again change. This is shown in Figure 7(b).
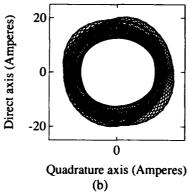

Figure 7. Direct-axis stator current vs. quadrature-axis stator current for (a) eccentric air gap and (b) eccentric air gap

Conclusion: This paper has presented a method for simulating an induction machine with a nonsinusoidal airgap flux distribution, in the presence of any arbitrary load conditions. This type of analysis is useful in analyzing machines with certain fault conditions or non-ideal operating conditions which cause a non-sinusoidal distribution of the windings or an eccentric airgap. A formulation was presented which describes the flux linkages in the machine in the presence of a time- and position-varying airgap length. Complete simulation results were presented which illustrate the harmonic components which exist in the stator current and torque produced by the machine as a result of the nonidealities.

References:

[1] H.C.Stanley, "An Analysis of the Induction Machine," *AIEE Trans.*, Vol. 57, pp. 751-757, 1938.

[2] P.C. Krause and C.H. Thomas, "Simulation of Symmetrical Induction Machinery," *IEEE Trans. on PAS*, Vol. PAS-84, No. 11, pp. 1038-1053, Nov. 1965.

[3] P.C. Krause, *Analysis of Electric Machinery*, McGraw-Hill, Inc., 1986.

[4] T.A. Lipo, *Analysis of Synchronous Machines, Course Notes for ECE 511*, University of Wisconsin, Madison, 1987.

FAILURE MECHANISMS FOR RIGID POLYURETHANE FOAMS USED IN FLEXURAL APPLICATIONS

Rick Kundrat
and Kenneth E. Hofer
L. J. Broutman & Associates, Ltd.
and Sunwoong Choi
Illinois Institute of Technology

Abstract: In structural sandwich applications where a panel is loaded perpendicular to its normal surface, the panel is introduced into a complex state of flexure. In this situation, the core-material will see the effect of tension on one side of the structure's neutral axis, and compression on the other side of the neutral axis. When such compound states of stress are induced, complex failure mechanisms may dominate.

Rigid polyurethane foam materials of several different densities were investigated in flexure with four point, one third span loading in an attempt to produce these compound states of stress in the materials. Failure mechanisms were observed and cataloged after testing. An effort was then made to define the failure modes based on foam densities and stress states at the time the crack/failure propagated through a given location in the foam sandwich core material.

The rigid foam failure modes should be useful in characterizing and tracking failures in more complex structures with skins and subject to unknown forces in the course of future failure analyses.

Introduction: Sandwich structures employing rigid polyurethane foam cores are finding their way into a greater variety of applications. Many factors have contributed to this fact including the development of improved fire retardant additives. Potential face sheet materials include wood, metal, neat plastic, and fiber reinforced thermosets. Increased structural strength to weight ratios, sound deadening, vibration damping, odor/vapor containment and fire/smoke suppression are some of the benefits or improvements gained by employing these types of sandwich structures.

One area where these types of sandwich structures have enjoyed significant penetration into the market is in the prefabricated building panel area, particularly in the area of prefabricated ceiling and roof panels. These panels are lighter than the materials they replace, are easier to install, and may be prefabricated in environmentally controlled factories.

In actual field-service applications these systems can be loaded perpendicular to their normal surface which puts them into a state of flexure. This stress state is then transferred through the face skins to the underlying core material. This then translates within the foam core to tension on one side of the neutral axis, and compression on the other side of the neutral axis.

The state of flexural stress in a homogeneous (i.e. non sandwich material) will produce a distinctive failure pattern. This pattern initiates on the tension side of the beam or panel and initially propagates in a plane perpendicular to the neutral plane of the structure. Eventually as the crack approaches the opposite (compressive side) of the beam or panel the crack turns producing the distinctive lip.

In brittle ceramic, glass or glassy plastics, the lip can be very pronounced with a sharp, razor edge or a final curve. In metals the lip is short with a strong mixture of extensive shear deformation.

In a sandwich panel or beam application, the skins generally carry the bulk of the tension and compressive stresses while the core assumes the beams shear and ties the two skins together thus playing the same roles in sandwich construction elements as the web and flanges in a steel I-beam.

Rigid foam core sandwich is a form of sandwich construction where the core is uniform and isotropic on a middle macroscopic basis. Furthermore the bonding of the core to the skins plays a key role in the overall integrity of sandwich structural elements subjected to flexure.

The nature of the propagation of cracks in rigid foam core under flexure is not as well understood as it might be but may be useful in establishing failure modes of complex structure in the same way that failure examination of glassy polymers, glasses, ceramics and metals is useful in understanding the progress and nature of failures in structures of those materials.

Test Program: Rigid polyurethane foam of different densities was selected for a test program aimed at ascertaining how would the failure patterns appear under flexural loading.

Four point flexural tests were performed on the rigid polyurethane foams to failure. The failure modes encountered were then examined and catalogued with regard to common characteristics.

Four blocks of polyurethane foam with densities 6.8, 10.0, 16.75 and 20.0 pounds per cubic foot (pcf) were obtained in 8" x 5" x 0.75" blocks. The rise direction was not designated on any of the sheets but uniformity in appearance on the cut faces was evident, thus obviating the need to include directionality as a variable. All sheets were sectioned to 1" widths. The 10 pcf and 20 pcf sheets were sectioned widthwise and the 6.8 and 16.75 pcf sheets were sectioned lengthwise. This produced two specimens; one of an 8" length and the other with a 5" length. To eliminate failure site as a variable, a shallow starter notch was introduced into the tension side of the specimens near the center of the specimen. This notch, the width of a razor blade, was cut across the width of the tension (lower) side of each specimen at its mid-length. Specimens were supported near the ends and loaded in four-point bending at one-third span load points (see Figures 1 and 2). Specimens were loaded to failure at a crosshead speed of 0.05 inches per minute.

Results: The ultimate stresses developed are shown in Table 1 below:

Table 1 - Foam Flexural Ultimate Stresses

Foam Density (PCF)	* 6.80	** 10.00	* 16.75	** 20.00	Solid 78.0
Failure Stress (PSI)	28.2	253.1	423.8	737.0	4000

* 8" long specimens   ** 5" long specimens

Solid polyurethane has a density of 78 pcf and an ultimate strength of 4000 psi.

Examination of the fracture surfaces after failure shows a commonality of appearance once the range of rigid foam densities is investigated. The beginning crack lies in a plane perpendicular to the centerline of the beam. The crack surface then turns increasingly away from this plane, i.e. the angle that the tangent plane makes with the beginning plane increases with distance from the starter tensile surface. In the final stage of fracture, the crack abruptly turns generally back toward the line across the compression or terminal surface which is directly opposite the starting crack on the opposite beam surface.

The less dense rigid foams had a fairly strong curvature of the failure surface. The densest foams had the least curvature.

In the final stages of fracture, direct examination of the fracture surface shows a rough appearance, typical of the surfaces where multiple potential crack paths exist, or bifurcation of the primary crack is taking place. Such multiple potential crack possibilities lead to an instability in the crack growth which in turn produces the rough appearing crack surfaces.

The higher density foams also exhibited slightly different behavior in that the relative percentage of the overall crack surface which has the rough appearance decreases as the foam density increases. Thus at the 6.8 lb/ft^3 density almost 30% of the fracture surface appears to be rough. At the higher density rigid foams the ratio of rough area to overall fracture area is less than 20%. The fracture surfaces are shown in Figures 3-6.

Conclusion: The fracture behavior of rigid polyurethane foams appears to be similar to the failure of brittle metals, ceramics and glasses, in that in its terminal stages a lip will be present at the compression surface of a flexed panel or beam.

The relative strengths of the foams in bending are shown in Figures 7 and 8. Figure 7 shows the foam strengths and the flexural strength of a solid homogeneous material plotted on the same chart. The behavior in the subscale and macroscale appears to be linear.

The failure behavior of virgin polymeric foam as cove in an aluminum or FRP skin sandwich requires future study. Certainly modification of the base behavior may occur. The fundamental path, fracture surface characteristics should be investigated for various shear bonding strengths to the skin materials. The authors plan to investigate such behavior modifying mechanisms in the future.

Density (pcf)	Length (L")
6.8	8.0
10.0	5.0
16.8	8.0
20.0	5.0

FIGURE 1. SPECIMEN DIMENSIONS.

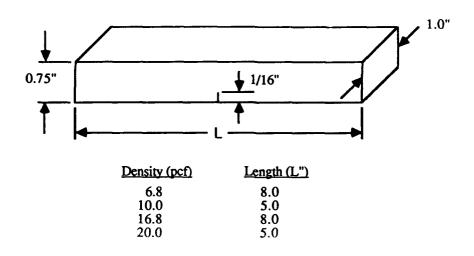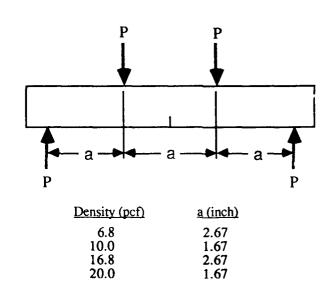| Density (pcf)	a (inch)
6.8	2.67
10.0	1.67
16.8	2.67
20.0	1.67

FIGURE 2. DIMENSIONS FOR FOUR POINT BEND TEST.

333

FIGURE 3. CRACK PROGRESS IN FLEXURE FOR 6.8 LBS/CU.FT. RIGID POLYURETHANE FOAM.

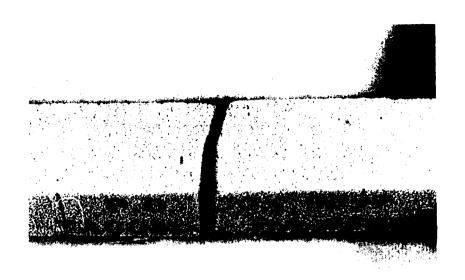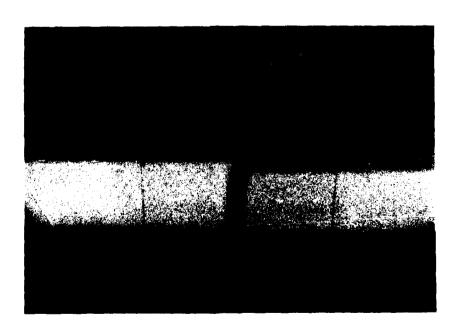**FIGURE 4. CRACK PROGRESS IN FLEXURE FOR 10 LBS/CU.FT. RIGID POLYURETHANE FOAM.**
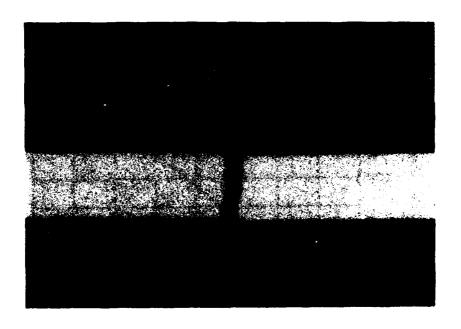
FIGURE 5. CRACK PROGRESS IN FLEXURE FOR 16.8 LBS/FT^3
RIGID POLYURETHANE FOAM.
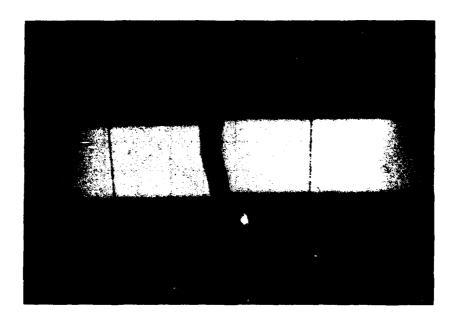
FIGURE 6. CRACK PROGRESS IN FLEXURE FOR 20 LBS/FT^3
RIGID POLYURETHANE FOAM.

335
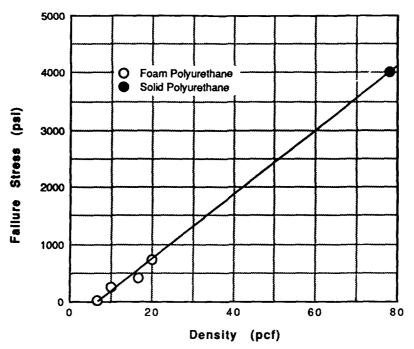
FIGURE 7. GRAPH OF FAILURE STRESS VS. DENSITY FOR
SOLID AND FOAM POLYURETHANES.
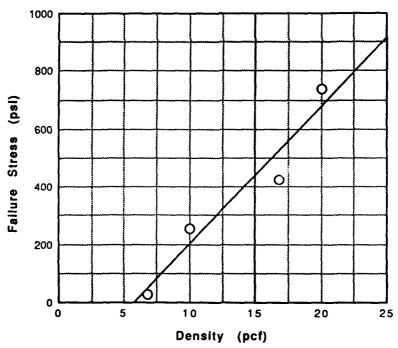
FIGURE 8. GRAPH OF FAILURE STRESS VS. DENSITY
FOR POLYURETHANE FOAMS.

CONDITION BASED MAINTENANCE
SYSTEMS ENGINEERING

Cochairmen: T. James DuBois
Southern California Edison Company

Robert L. Tregoning
Naval Surface Warfare Center

SHIPBOARD MAINTENANCE: A "RE-TOOLING" PROCESS

Anthony M. Cieri
Naval Surface Warfare Center
Carderock Division
Ship Systems Engineering Station
NAVSSES (Code 10)
Philadelphia Naval Base
Building 633
Philadelphia, PA 19112-5083

Augustine DiGiovanni
Naval Surface Warfare Center
Carderock Division
Ship Systems Engineering Station
NAVSSES (Code 101)
Philadelphia Naval Base
Building 633
Philadelphia, PA 19112-5083

ABSTRACT: The realization that there will be fewer assets and reduced budgets as the Navy progresses into the 1990s, coupled with the continuing increase in ship's maintenance costs, requires a fundamental change in the Naval maintenance community. The Condition Based Maintenance Branch at the Naval Ship Systems Engineering Station (NAVSSES), Carderock Division, Naval Surface Warfare Center (NSWC) provides engineering and integration support to the Navy for shipboard equipment assessment. This paper will examine some of the limitations of the existing process and discuss development efforts currently underway to effectively integrate current monitoring/analysis technology and techniques. The paper will offer an engineered approach to provide shipboard condition assessment by integrating oil sampling data, vibration data, equipment operational data and other equipment/system condition data.

KEY WORDS: Automated diagnostics; condition based maintenance; expert systems; oil analysis; shipboard maintenance program; vibration monitoring.

INTRODUCTION: The pumps, motors, electrical equipment and other pieces of machinery installed onboard Navy ships are currently monitored on a regular basis by routine watchstanding personnel. Operating parameters are manually recorded on handwritten logs that are subsequently reviewed and filed. This methodology has many shortfalls, most notably the subjective nature of the log review process and the lack of access to historical data. NSWC Carderock Division is striving to equip shipboard personnel with modern performance monitoring and data collection tools that will provide the Navy with an affordable approach to equipment assessment. The envisioned shipboard systems will have computer-based automated programs that will sort and trend data, highlight abnormal readings and focus ship's force towards generating work requests. Vibration, performance and oil sampling data for selected equipment will be manually collected by shipboard personnel equipped with hand held data collectors and fed into the maintenance data collection program. This system will streamline the shipboard maintenance planning process, improve the operational readiness of our ships and conserve maintenance dollars. The Navy faces the challenge of developing, testing and implementing fleetwide shipboard diagnostic systems that integrate the different types of data collection into a coherent maintenance system. During the design phase, the technical aspects of transferring new technologies, tools and processes to the shipboard maintenance infrastructure must be addressed.

Once designed, the shipboard systems will be installed on proof of concept prototypes. Lastly, lessons learned from the test and evaluation of the prototypes will be incorporated and shipboard diagnostic systems will be installed fleetwide.

CONDITION BASED MAINTENANCE: In 1988, the Chief of Naval Operations directed an overall maintenance strategy that is based on the principles of Reliability-Centered Maintenance (RCM).[1]   Reliability-Centered Maintenance, by its nature, is intended to prevent failures. RCM is a methodology to develop preventive (failure precluding) maintenance tasks. They can be time based, on condition or failure finding. Equipment repairs are made to correct current faults, preempt further degradation and prevent future failures.

The predominant maintenance philosophies currently in use are fix-when-fail and time-based. Fuses are maintained under the fix-when-fail philosophy; they are replaced when they blow. In general, the fix-when-fail philosophy is applied to equipment that is low in cost, difficult to trend, and/or simple to repair. Time-based maintenance calls for replacement of components or completion of equipment overhauls at fixed time frequencies (calendar or operating hours). Time-based maintenance is used for components, equipment and systems ranging in complexity from oil filters to gas turbines. The newest maintenance philosophy currently in use, predictive maintenance, has great potential for improving equipment availability while conserving maintenance funds. Predictive maintenance is based on the premise that equipment condition can be assessed on a periodic or ongoing basis by comparing actual performance data to a set of desired specifications. Properly implemented, predictive maintenance gives maintenance personnel a better picture of equipment condition. Minor equipment flaws are identified before they lead to major failures and maintenance funds are targeted for equipment in actual need of repair. Insurance repairs, which are driven by uncertainty as to equipment condition, will no longer be necessary. In turn, the infant mortality failures often associated with major repairs/overhauls will become less prevalent.

Condition based maintenance (CBM) is comprised of elements of all three of these maintenance philosophies. The key to successful implementation of CBM is application of the proper level of monitoring, evaluation and trending for each piece of equipment. Predictive maintenance is not appropriate and/or cost effective in all cases. Today's maintenance managers face the challenge of creating the proper mix of fix-when-fail, time-based maintenance, and predictive maintenance. The principles of RCM, coupled with integrated diagnostic tools and techniques, when implemented correctly, will aid the solution to this challenge.

THE END PRODUCT

The foundation of CBM and surface ship maintenance as a whole is envisioned as a shipboard computer based maintenance system that provides greater capability to the ship for condition assessment. Complex equipment, such as propulsion boilers and steam turbines, will have on-line sensors to allow continuous on-line monitoring of key performance

parameters. Other equipment, such as pumps, motors, and electronics equipment will be monitored on a periodic basis with hand held data collectors. In both cases, performance, vibration and other diagnostic data will be fed into a computer-based automated diagnostic (expert) system that will maintain and trend the data, highlight abnormal readings, and recommend minor repairs (alignment, bearing replacement, etc.) and system grooming when required. The maintenance system will also provide monitoring and maintenance training, technical manual and logistics information and will be integrated with the 3M and supply systems. Outputs from the system will include material management and administration, work definition, logistics support, and measures of effectiveness.

The shipboard maintenance system will be linked to, and supported by, the shoreside maintenance activities. Satellite communication links will feed detailed information to Port Engineers and other maintenance managers upon demand and macro level information (availability/MTBF and repair costs) will be fed to a centralized database to facilitate comparison of equipment performance at the equipment, system, platform, hull, and fleet levels. This will provide better and more real time condition information for shoreside planners to improve business decisions regarding completion of major repairs, availability scheduling, equipment alterations, etc. Performance Monitoring Teams (PMTs) will conduct ship visits prior to major availabilities to review the ship's performance data, train ship's force in grooming, diagnostics monitoring and analysis techniques, and conduct more sophisticated performance evaluations requiring high cost test equipment or specialized training. Condition-based repair recommendations will be automatically generated and forwarded to availability planners and the Type Commanders (TYCOMs). Class Maintenance Plans will be updated to mandate that repairs be scheduled based on actual equipment performance and condition. The PMTs will conduct post-availability ship visits to re-baseline repaired equipment and to target equipment and systems requiring repair during the next operating cycle. Specialized maintenance programs (diesel inspections, boiler inspections, etc.) will not be eliminated. They will be provided the tools required to implement CBM and will be benefit from the standardization of procedures, analysis, and diagnostics that will occur as the Navy evolves to CBM. Responsibilities will be clearly defined and redundancies of effort will be eliminated. In essence, a continuum of maintenance will be established as the feedback loop will be closed. Cost and availability data will identify areas of concern. Engineering reviews will be conducted by In-Service Engineering Agents (ISEAs) and Life Cycle Managers (LCMs) and design/logistical shortcomings will be resolved. The condition assessment system will be structured as shown in Figure 1. Our ships will be maintained as efficiently as possible and sound business principles will be the foundation of the entire structure.

THE AEC PROGRAM: The Assessment of Equipment Condition (AEC) Program assists the Type Commanders in work definition and availability planning. The objectives of the program are to provide better work definition through condition-based maintenance and to provide the impetus of fleetwide implementation of CBM. Non-intrusive evaluations of shipboard equipment are made by Performance Monitoring Teams (PMTs), which conduct two ship visits per operating cycle (pre and post depot level
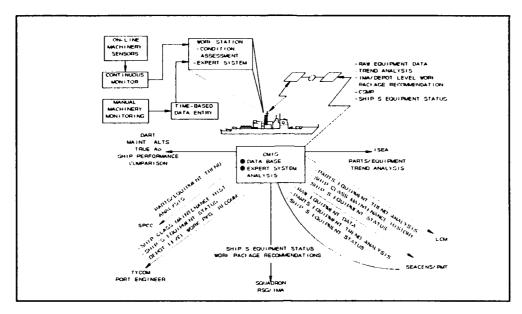
341

Figure 1. Condition Assessment System

availability). Performance data is collected and forwarded to the appropriate Naval Sea Support Center (NAVSEACEN) for review. The NAVSEACEN uses this data to make repair recommendations and to recommend deferrals for scheduled maintenance actions deemed to be unnecessary. The AEC Program is also heavily involved in several of the CBM prototypes listed above.

Prior to 1989, the AEC Program concentrated its efforts on a select number of ship classes (FF-1052, CG-47, DD-963 and DDG-993). PMTs conducted quarterly ship visits to monitor equipment performance. The primary objectives of the AEC Program were to conduct condition based maintenance in an effort to extend the time between depot level availabilities and to ascertain the applicability of condition based maintenance methodology to shipboard systems. As a proof of concept, the AEC Program proved to be very successful. It supported the extension of depot level availabilities and built an impressive record in reducing maintenance costs.

AEC Expansion: In the latter part of 1989, the AEC Program was directed by the Naval Sea Systems Command (NAVSEA) to expand its coverage to all surface ships and the entire platform and to define repairs for all high maintenance burden systems prior to depot level availabilities. As a first step, the AEC Program conducted a detailed cost analysis to identify high maintenance cost systems and equipment. A list of candidate systems and ship classes to be covered was then developed. Approval of this list was obtained from the Type Commanders and NAVSEA. Maintenance Requirement Cards (MRCs) and analysis guides for 34 HM&E and combat weapons systems were developed during fiscal year 1992 and development of an additional 13 will be completed in fiscal year 1993. The MRCs are being forwarded to the commodity specialists for 3M issue and the analysis guides are being promulgated to the PMTs and the NAVSEACENs. Advance copies of the MRCs

342

are also sent to the PMTs and NAVSEACENs so that shipboard equipment can be assessed pending issue of the next Semi-Annual Force Revision from 3M. The AEC Program has initiated a plan for implementing the MRCs developed in fiscal years 1992 and 1993. This plan covers systems training for PMT personnel, procurement of required Support and Test Equipment (S&TE) and fulfillment of PMT manning requirements and is expected to be complete for HM&E systems by the end of fiscal year 1994.

AEC and CBM: The AEC Program's history and recent expansion have been described in this paper as both are pertinent to ongoing efforts to develop shipboard diagnostic systems. The lessons learned by AEC program managers at the Naval Ship Systems Engineering Station (NAVSSES) as the program evolved in the 1980's are being used to ensure that appropriate shipboard equipment is targeted for monitoring and evaluation. Additionally, in many cases, the AEC procedures and analysis guides developed since the latter part of 1989 will be used as a starting point for developing automated diagnostic software. NAVSSES will draw heavily on AEC experience as the initiative described in the ensuing paragraphs is executed.

EDMS INITIATIVE: The Engineering Data Management System (EDMS) is a computer based shipboard maintenance system being developed by NSWC Carderock Division. The first phase of EDMS development will not involve shipboard equipment monitored by on-line sensors. This limitation was imposed to conserve funds, to avoid redundant efforts, and to focus on data collector and diagnostic software development. EDMS is being initially designed to receive equipment data inputs, provide alarms and to trend data into a usable machinery performance history.

Parameters Evaluated: The equipment data inputs for EDMS will include machinery vibration data, oil analysis results and equipment performance data. These inputs will be used to assess and trend the condition of selected shipboard equipment and to locate and identify faults resulting from excessive operating conditions, poor or improper lubrication, improper maintenance/repair and operator training deficiencies.

Vibration monitoring will be used initially to provide warnings of deteriorating equipment performance and, eventually, to identify misalignment and imbalance in rotating and reciprocating machinery, and deteriorating or defective bearings and gears. On a periodic basis, ships force personnel will download a vibration survey route from EDMS to an Advanced Vibration Meter (AVM). Vibration data will then be collected and fed back into the system which will trend and maintain historical vibration data for each piece of machinery monitored and compare vibration levels to similar equipments onboard. The AVM will indicate an alarm condition when vibration inputs for a piece of equipment exceed established alert levels and the system will log all alerted machines and generate a daily vibration alert report. Initially, only broadband vibration levels will be collected and trended. Historic data shows that roughly 40 percent of alerted machines are actually in need of repair. Accordingly, ship's force will track alerted machines and, when warranted by sudden increases or increasing trends, troubleshoot the machine in

343

question or request a PMT narrowband vibration cut. Once ship's force
personnel have become familiar with EDMS, the diagnostic capabilities of
the AVM will be enabled so that ship's force can test for misalignment,
imbalance or bearing wear. This added capability will have a twofold
benefit in that bearing faults not discernible in broadband readings will
be identified and machines with unusual or subtle problems will be
targeted for PMT testing.

Oil analysis, like vibration monitoring, provides valuable insight to
equipment health. EDMS will cue ship's force to perform four periodic
tests on shipboard lubricants. These tests for wear particles, viscosity,
water content, TBN, fuel dilution, and particulate contamination will
provide a clear picture of the condition of the ship's lubricants and
allow ship's force to monitor for excessive wear and deterioration of
equipment parts. The results of these tests will be uploaded to EDMS for
comparison with desired specifications. When warranted, a NOAP sample
will be called for. Placing this capability onboard our ships can yield
significant benefits in three ways. It will reduce the time required to
identify lubricant problems and, in some cases, prevent equipment
failures. It can also reduce the workload of NOAP laboratories which are
currently overloaded with routine samples. This reduction in workload, in
turn, would enable the NOAP laboratories to spend more time evaluating
lubricant samples taken from equipment with actual problems.

Finally, operational and performance data will be collected to evaluate
pumps, motors, compressors, engines and heat exchangers. Initially, the
data collected will be limited to the operational data currently monitored
by shipboard watchstanders. As the development of knowledge based
algorithms is completed, performance tests will be added to the system.
These tests will include flow analyses, alignments and load tests and will
be used to assess the conditon of shipboard equipment. In many cases, the
data collected will be identical to the data currently used by the AEC
Program to provide repair recommendations to the Type Commanders.

Collectively, the vibration surveys, oil analyses and performance tests
will be the key element of the end product described in this paper as they
will give ship's force a clear picture of the actual condition of
shipboard equipment and allow intelligent maintenance decisions to be
made.

Computer Assets: EDMS will reside on a 486DX IBM compatible computer.
The computer will have a 10 minute uninterrupted power supply to protect
valuable information and a modem to allow for future ship to shore
communications. The system's software will contain the algorithms
(knowledge based systems) that will store performance, vibration and oil
data and transform the same into equipment condition information. The
initial algorithms will maintain, trend and alarm data as described above.
The information and algorithms will serve as the foundation for the more
sophisiticated capabilities yet to be developed and/or added.

Initial Capabilities: Initially, EDMS will cue ship's force to collect vibration, oil and operational data, download data collecting routes, provide alarm functions, generate logs and 8 O'Clock reports, maintain historical data files, and perform data trending/comparison functions. The system will also generate hard copy 4/90/2K's. These capabilities will start the Navy wide evolution from paper and pen to hand held loggers and computers and provide the foundation for planned enhancements.

Future Enhancements: Incorporating "Knowledge Base Engineering" will take EDMS to the next level, in that performance data, in conjunction with fault isolation, will provide troubleshooting assistance and assist ship's force in its ability to effect repairs and make maintenance recommendations. Fully developed, the system will make repair recommendations, store inventory/configuration data, schedule equipment maintenance and monitoring, maintain parts inventories, manage cost and performance parameter databases and generate work requests and other required reports. The envisioned system is shown in Figure 2.

Figure 2. The Engineering Data Management System
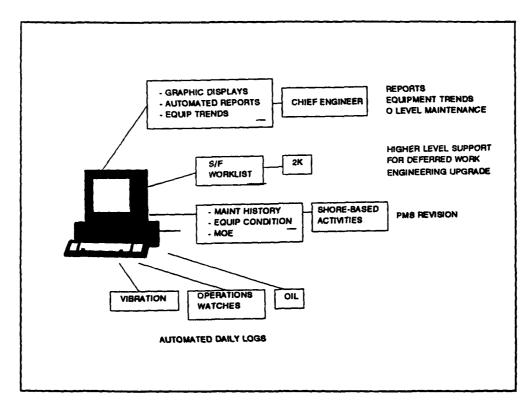
EDMS Prototypes: Four proof of concept EDMS prototypes are envisioned; all on the East Coast. They will be installed on USS KIDD (DDG-993), USS SCOTT (DDG-995), and two FFG-7 class platforms. Development, procurement, installation and testing of these prototypes will be coordinated by NSWS Carderock Division under the sponsorship of the Commander, Naval Surface

345

Forces, Atlantic Fleet (COMNAVSURFLANT). The design of EDMS is complete and most of the computer hardware/software, data collectors and test equipment required to build the four prototypes have been developed. In many cases, the components have also been tested and are in use for other applications. When the equipment is on hand, NSWC Carderock personnel will install EDMS on the four designated platforms, provide logistics support, indoctrinate ship's force personnel and initiate the test and evaluation process. EDMS will be installed in two phases.

Initial Installation: The EDMS hardware will be installed onboard the four selected platforms and ship's force will be introduced to the concepts of vibration surveys, shipboard oil analysis and hand-held data collectors. The primary goals during this phase are to familiarize ship's force with EDMS and to provide the ship with the initial tools to perform their own maintenance planning.

The EDMS library of knowledge based algorithms will be expanded to cover the systems listed in Table 1. Additionally, the data loggers will be incorporated into watch standers' normal rounds in machinery spaces and training/logistics features, such as technical manual extracts, Engineering Operating Sequencing System (EOSS) procedures and Maintenance Requirement Cards (MRCs) will be loaded into the EDMS computer. PMS and EOSS programs are currently being automated to the point where the Fleet

SYSTEM	VIB	OIL	AUTO-LOG PERFORMANCE	ON-LINE CAPABILITY
HP Air	X	X	X	X
LP AIr	X	X	X	X
Dry Air			X	
Firemain System	X		X	*
Seawater Service	X		X	
Air Conditioning	X	X	X	X
400Hz Pwr Supply			X	*
Steering	X		X	*
Main Propulsion Train	*	X	X	X
Distillation	X	X	X	X
Fuel Oil Service	X		X	X
Fuel Oil Fill & Xfr	X		X	*
Ship Service Generators	X	X	X	X
Drain/Ballast/Trim	X		X	*
Deck Machinery	*	X	X	X
Boat Handling/Stowage		X	X	
Main Propulsion Lube Oil	X	X	X	X
L.O. Fill/Xfr/Purif	X		X	
CHT System	*		X	*
Oily Water Waste System			X	
Fresh Water	X		X	X
Degaussing			X	
Cathodic Protection				X

X - Technologies available and can be used onboard Naval ships
* - Technologies developed but not used onboard Naval ships

Table 3. EDMS Systems

will begin receiving CD ROMs with their hard copy PMS and EOSS issues during Fiscal Year 1993. Also, technical manuals are being digitized for easy storage and distribution. Linking these automation efforts to EDMS will significantly improve the efficiency onboard the EDMS platforms as maintenance personnel will be able to print out technical manual drawings and take them to the maintenance cite vice checking out bulky manuals and subjecting them to the hazards of the shipboard environment. Additionally, the control of manuals, MRCs etc. will be greatly simplified as inventory and change information will be stored in the EDMS computer rather than on handwritten books. and Configuration Change Requests (4790/CKs) will identify required changes to technical publications.

T&E/Implementation: A test and evaluation (T&E) plan will be used to evaluate the accuracy and reliability of each EDMS prototype and to determine if the system is user friendly and effective. The plan will also be used by shipboard personnel as a training tool. Measures of Effectiveness (MOE) will objectively monitor and trend the system's return on investment (ROI); both in real dollars and in terms of material readiness. After a 6 to 12 month evaluation period, a final report will be issued for each prototype. These reports will contain cost benefit analyses, risk management assessments and recommendations concerning the applicability and effectiveness of the tools and techniques applied. With these reports in hand, along with feedback received from interviews with shipboard personnel and similar reports from the M-CAS prototypes, NAVSSES will be able to build the integrated shipboard diagnostic system and proceed to Navy-wide implementation.

MACHALTs/SHIPALTs: Machinery Alterations (MACHALTs)s are used by the U.S. Navy to effect changes to equipment and systems where the changes are contained within the boundaries of the individual equipment or system and have limited impact on other (external) equipment or systems. A MACHALT is defined as a planned change, modification or alteration to any equipment in service (shipboard or shore based) when it has been determined that the alteration or modification can be accomplished without changing an interface external to the equipment or system; is a modification made within the equipment boundary or is a direct replacement of the original equipment design; can be accomplished without the ship being in an industrial activity; and will be accomplished individually and not conjunctive with a SHIPALT or other MACHALT.[2] The MACHALT Program employs a kit installation concept (Figure 3) that enables equipment changes to be accomplished in an expeditious manner and eliminates them from the formal Ship Alteration (SHIPALT) process. The program has been so successful that NAVSSES managers now use the MACHALT process to manage SHIPALTs as well.

CONCLUSIONS: As the Navy approaches the 21st Century, it must learn to do more for less. Condition Based Maintenance (CBM) is the chosen means for getting ship maintenance costs under control. To successfully transition to CBM, the Navy must develop and install shipboard maintenance programs to make our fleet units more self sufficient and capable of making accurate repair recommendations. The Navy faces the challenge of developing, testing and implementing this shipboard system and changing
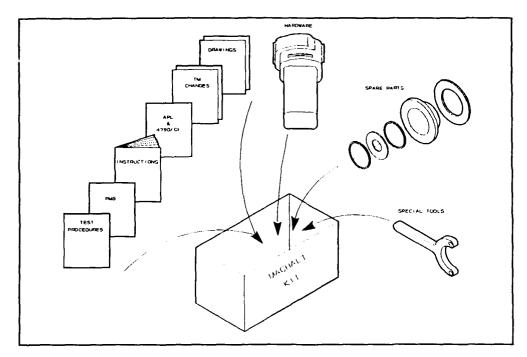
347

Figure 3. MACHALT Kit Concept

the maintenance infrastructure to support a ship oriented maintenance hierarchy. The Engineering Data Management System (EDMS) is being developed by NAVSSES to serve as the foundation of the shipboard maintenance program. Fully developed, EDMS will enable ship's force personnel to feed performance, vibration and oil data fed into a computer-based automated diagnostic system that will maintain and trend the data, highlight abnormal readings, and recommend minor repairs (alignment, bearing replacement, etc.) or system grooming when required. EDMS will also provide monitoring and maintenance training, technical manual and logistics information and will be integrated with the 3M and supply systems. Outputs from the system will include material management and administration, work definition, logistics support, and measures of effectiveness. The transition to CBM is predicated on the successful development, testing and implementation of systems such as EDMS.

REFERENCES

1. Chief of Naval Operations, OPNAVINST 4700.7H, "Policies and Procedures for Maintenance of Ships," 30 October 1987.

2. Commander, Naval Sea System Command, NAVSEAINST 4720, "Policy and Procedures for Implementation of Machinery Alterations (MACHALTS) on HM&E Equipment and Systems," 23 March 1983.

CONDITION MONITORING AND DIAGNOSTICS OF A RECIPROCATING COMPRESSOR

Daniel J. McCarthy
Research Assistant, Mechanical Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

H. A. Gaberson, Ph.D.
Senior Technologist
Naval Civil Engineering Laboratory, Code L72
Port Hueneme, CA 93043-5003

Abstract: Vibrational measurements made on the casing of a machine contain information that can be exploited for diagnostic purposes if the signals are processed properly. One approach that has been studied in recent years is waveform recovery, whereby the available vibration signal is processed to obtain information regarding the forces that caused the vibration. A current research project at MIT is aimed at furthering the basic knowledge required for the recovery of impulsive source waveforms for use in diagnosing the faults of reciprocating machinery. In this paper we demonstrate an improved technique that combines the use of cepstral smoothing and minimum-phase decomposition. In addition, we introduce the use of a time-frequency domain technique, the "short-time coherence," that can helpful for use in determining the times and frequency ranges over which to perform the inverse filtering.

Introduction: The manner in which machines vibrate contains information about their operating condition; indeed, many signal processing techniques exist for exploiting the vibration signatures of rotating machinery for diagnostic purposes. However, these same techniques provide little information for diagnosing the condition of reciprocating machinery because, in addition to narrowband

349

vibrations due to rotating components, there are large amplitude broadband vibrations due to such events as valve impacts and sharp variations in the pressure waveforms within the cylinders and manifolds. Knowledge of the timing and strength of these vibration-generating events can be useful for diagnostics[1]; unfortunately, practical considerations dictate that the vibrational measurements must be made on the casing of the machine, where the source signal has become contaminated by dispersion, reverberation, multi-path transmission, and overlapping of the vibrations due to various events.

In the first part of this paper, we demonstrate the recovery of known impulsive source waveforms using cepstral smoothing and minimum-phase decomposition. In the second part, we describe the compressor used in the research. Finally, in part three we introduce the use of a "short-time coherence" to determine the time and frequency ranges over which to perform the inverse filtering, and its usefulness for separating simultaneous or closely occurring source events.

Part I. Inverse Filtering: A model for vibration transmission through a linear, time-invariant system is given by:

$$Y(f) = X(f)H(f) \tag{1}$$

where $X(f)$ and $Y(f)$ are Fourier transforms of the excitation input and vibrational response, respectively, and $H(f)$ is the transfer function describing the vibration transmission properties of the system. This model relies an accurate estimate of the transfer function, which can then be used to work back from a measurement of the response to estimate the excitation. This is done by multiplying the measured output by the inverse of the transfer function estimate:

$$X(f) = Y(f)/H(f) = Y(f)H(f)^{-1} \tag{2}$$

Unfortunately, the assumptions of linearity and time-invariance are often violated in practice because vibration transmission through a machine is affected by variations in operating characteristics, such as temperature and load. Additional variability is introduced in diagnostics because the sensors are often not permanently mounted and will therefore vary in location. Also, for this diagnostic technique to be widely accepted, it must not require that transfer functions be measured for every machine to be monitored–a crude estimate from a nominally identical machine structure must suffice. Figure 1 shows typical transfer function variations due to changes in temperature as measured on our test compressor. The plots show up to a 30 dB variation in magnitude and approximately 4π in phase for a temperature range from $75°F$ to $210°F$, and a frequency range from DC to 2 kHz. The 2π jumps in phase can be attributed to zeros of the transfer function which drift back and forth between minimum and non-minimum phase behavior. Non-minimum phase zeros in a transfer function are particularly troublesome because they invert to unstable poles of the inverse filter, which cannot then be stable and causal[2].

Previous researchers, using inverse filtering in conjunction with cepstral smoothing to reduce path variability, have been able to recover information about the timing

vibrations due to rotating components, there are large amplitude broadband vibrations due to such events as valve impacts and sharp variations in the pressure waveforms within the cylinders and manifolds. Knowledge of the timing and strength of these vibration-generating events can be useful for diagnostics[1]; unfortunately, practical considerations dictate that the vibrational measurements must be made on the casing of the machine, where the source signal has become contaminated by dispersion, reverberation, multi-path transmission, and overlapping of the vibrations due to various events.

In the first part of this paper, we demonstrate the recovery of known impulsive source waveforms using cepstral smoothing and minimum-phase decomposition. In the second part, we describe the compressor used in the research. Finally, in part three we introduce the use of a "short-time coherence" to determine the time and frequency ranges over which to perform the inverse filtering, and its usefulness for separating simultaneous or closely occurring source events.

Part I. Inverse Filtering: A model for vibration transmission through a linear, time-invariant system is given by:

$$Y(f) = X(f)H(f) \tag{1}$$

where $X(f)$ and $Y(f)$ are Fourier transforms of the excitation input and vibrational response, respectively, and $H(f)$ is the transfer function describing the vibration transmission properties of the system. This model relies an accurate estimate of the transfer function, which can then be used to work back from a measurement of the response to estimate the excitation. This is done by multiplying the measured output by the inverse of the transfer function estimate:

$$X(f) = Y(f)/H(f) = Y(f)H(f)^{-1} \tag{2}$$

Unfortunately, the assumptions of linearity and time-invariance are often violated in practice because vibration transmission through a machine is affected by variations in operating characteristics, such as temperature and load. Additional variability is introduced in diagnostics because the sensors are often not permanently mounted and will therefore vary in location. Also, for this diagnostic technique to be widely accepted, it must not require that transfer functions be measured for every machine to be monitored–a crude estimate from a nominally identical machine structure must suffice. Figure 1 shows typical transfer function variations due to changes in temperature as measured on our test compressor. The plots show up to a 30 dB variation in magnitude and approximately 4π in phase for a temperature range from $75°F$ to $210°F$, and a frequency range from DC to 2 kHz. The 2π jumps in phase can be attributed to zeros of the transfer function which drift back and forth between minimum and non-minimum phase behavior. Non-minimum phase zeros in a transfer function are particularly troublesome because they invert to unstable poles of the inverse filter, which cannot then be stable and causal[2].

Previous researchers, using inverse filtering in conjunction with cepstral smoothing to reduce path variability, have been able to recover information about the timing
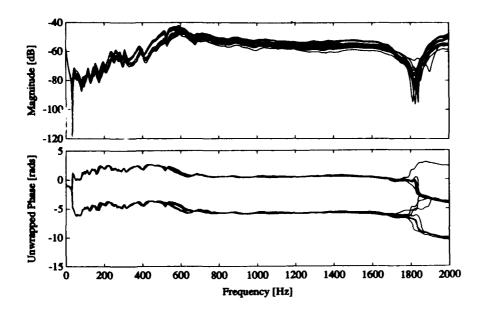
350

Figure 1: Variations in transfer function magnitude and phase due to changes in temperature ($75°F$ to $210°F$).

and strength of pressure waveforms in diesel engines, and have displayed their usefulness for diagnostics[3]. Additional research has led to the development of a technique that uses minimum-phase decomposition for recovering impulsive source waveforms in rooms[4]. We have found that we get the best results by combining these two techniques. We first perform the minimum-phase decomposition[5], so that the linear phase can be obtained as the average group delay of the all-pass part of the signal, and then we cepstrally-smooth our signals using a homomorphic deconvolution approach[6]. To get reliable linear phase information, light exponential windowing must be applied to the signals used in estimation of the transfer function, as well as to the response signal that is used in the recovery process. Additionally, the range over which the group delay averaging should be performed will vary–the best results are obtained by not including frequency regions surrounding the "drifting zeros" which were described in conjunction with Figure 1. The processing is applied to both the response signal and an estimate of the transfer function to obtain $\tilde{Y}_{min}(f)$ and $\tilde{H}_{min}(f)$. The recovered signal $x_r(t)$ is then given by:

$$x_r(t) = F^{-1}\left[\tilde{Y}_{min}(f)\tilde{H}_{min}(f)^{-1}\right] \tag{3}$$

As a proof-of-concept we have experimented with recovery of impulsive source waveforms applied to the compressor structure using a hammer instrumented with a load cell for measuring the applied force. The transfer function (see Figure 1) was obtained at $195°F$ and then used to recover source waveforms with the structure at $110°F$. Figure 2a shows the inadequate result obtained with basic inverse-filtering, while Figure 2b shows the accuracy of the recovery when the extra processing is performed. The transfer function used in this case is one that we
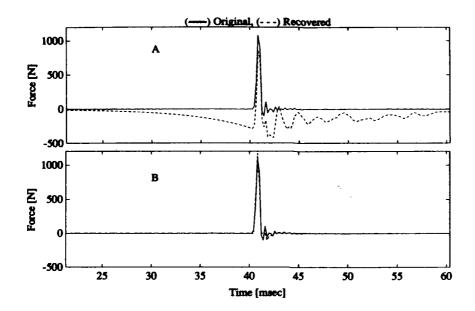
351

Figure 2: Recovery of impulsive source waveforms applied to the compressor structure using an instrumented hammer: A. Basic inverse filtering; B. Inverse filtering in conjunction with minimum-phase processing and cepstral-smoothing.

expect to use for recovery of valve impact waveforms in the operating compressor.

Part 2. Compressor Description: The machine that we are using for our experiments is an Ingersol-Rand Type 40, two-stage, air-cooled air compressor like that shown in Figure 3. The compressor is belt-driven by a 40 HP electric motor rated at 1765 RPM; the compressor itself is rated at 870 RPM for a load of 125 psi. The first stage of compression is accomplished by two 7.5 inch diameter pistons, the second with a single 6.25 inch diameter piston, while all three have a 5 inch stroke. The compressor is attached to a storage tank equipped with a valve that permits loading of the compressor at constant pressures from 0 to 120 psi.

The compressor valves are of the reed type, which allow air to flow one way but not the other. Each valve consists of five to seven leaf-spring/channel combinations that open and close individually, but in unison, as shown in Figure 4. A common source of failure, the valves have been the focus of our research to date. The valves are instrumented with strain gages to record times of opening and closing, as well as with accelerometers to provide information about the strength of the valve impacts against the valve body.

As a compressor operates, there are an abundance of vibration-generating sources, broadband as well as narrowband. The most distinct sound one hears when a compressor is running is the plosive "-pa-pa-pa-" that occurs when the valves open and close. We are in the process of trying to determine if, as with diesel engines[1], the "-pa-" sound is due to sharp variations of pressure in the cylinders and
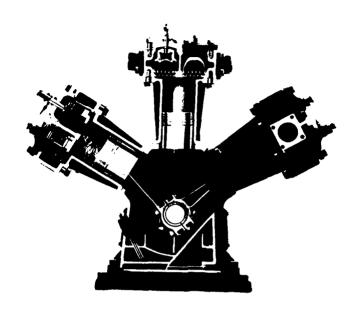
352

Figure 3: Cutaway view of the compressor.
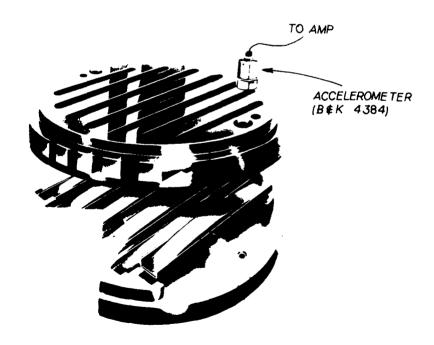

TO AMP

ACCELEROMETER
(B&K 4384)

Figure 4: Valve mechanism.

manifolds when the valves open and close. Another event, simultaneous with the pressure discontinuities, but that is not as easily heard with the "naked" ear, is the impact of the individual valve channels against the valve body when the valves open and close. Narrowband vibrations due to rotational components also show up in our vibrational measurements, albeit at a much lower level (approximately 3 g's peak-to-peak for a filter cutoff of 13.2 kHz) than we measure for the transient vibrations (100 g's peak-to-peak on the valve body and approximately 15 g's peak-to-peak for the casing acceleration for a filter cutoff of 13.2 kHz).

Part 3. Short-Time Coherence: The task of source waveform recovery is made more difficult when multiple sources of vibration are present. We have attempted to recover valve impact forces in the compressor without regards to the effects of simultaneously occurring events, chiefly the pressure discontinuities, but have found that the recovered waveforms do not correlate well with the measured acceleration on the valve body. Further processing is required to separate the events by frequency content. One practical way to discriminate against unwanted sources is to make the response measurement as close as possible to the source that you are interested in, with the added benefit of reducing transfer function variability. Unfortunately, the source events that interest us in the compressor occur nearly simultaneously in time and have very little spatial separation.

Another method that we are finding useful for discriminating between sources is based on a time-frequency domain technique derived from the short-time fourier transform (STFT), which we refer to as the short-time coherence (STC). The STC can be used to determine the time and frequency ranges over which the sources of vibrational energy are strongly coherent with the measured vibrations, but not with each other. Once this is determined, the waveform recovery can be performed on vibration data that has been filtered in the frequency range, or ranges, of high coherence. We have found that, when the inputs to the system are impulsive in nature, we can adequately recover information regardi ⌐ the strength of the input, given that the filtering is not too severe. Further wor⌐ ⌐s necessary to understand the filtering effects on the timing information.

By examining the STFTs of the discharge valve body acceleration, and of the pressure in the second-stage cylinder, as shown in Figures 5 and 6, respectively, we can see that there are broad-band events that occur simultaneously in each when the valves open and close. The STFTs shown in this paper cover the frequency range from DC to 13.2 kHz for just under two machine cycles. The STFT of the valve impacts, for which only the highest level contours have been plotted, shows that there is significant energy over the entire frequency range from DC to 13.2 kHz after impact. In fact, we have examined the valve plate acceleration out to 40 kHz without seeing any apparent fall off, which is partly due to the fact that acceleration increases with the square of frequency, though there is clearly abundant excitation at these high frequencies due to the sharpness of the impact forces. The STFT of the pressure waveform in Figure 6 shows an increase in broadband energy due to pressure discontinuities when the valves open and close,
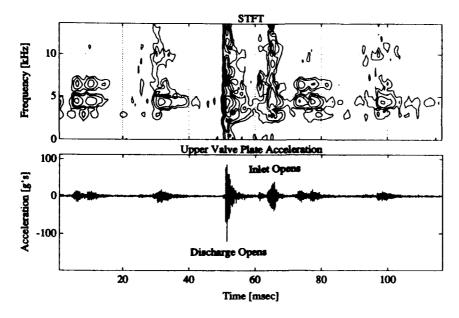
354

Figure 5: STFT and time-waveform of the valve body acceleration.

with the highest levels occurring below 1500 Hz. The STFT of the casing acceleration, shown in Figure 7, displays strongly transient characteristics similar to those of the valve acceleration. Figures 5, 6, and 7 also show the time-waveforms of the valve impact, cylinder pressure, and casing acceleration, respectively, with some individual events labeled.

The most basic form of the coherence function relating an input $x(t)$ and an output $y(t)$ of a linear system is the ordinary coherence function given by:

$$\gamma_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} \tag{4}$$

where $G_{xy}(f)$ is the one-sided *cross-spectral density*, and $G_{xx}(f)$ and $G_{yy}(f)$ are the one-sided *power spectral densities* of the two time records. For a linear system with incoherent inputs, the coherence function can be interpreted as the fractional portion of the mean square value at the output $y(t)$ that is contributed by the input $x(t)$ at frequency f.[7]

When dealing with a tightly-coupled mechanical system like a compressor, where there are many sources present, care must be taken in intrepreting the ordinary coherence function, because the existance of sources that are coherent with each other can lead to erroneously high levels of coherence between inputs and outputs. We can minimize this problem by seeking regions in the time-frequency domain where a single source is highly coherent with the output and is incoherent with the other sources. In the analysis to be described, we have used the acceleration of the upper valve plate as a measure of the strength of the valve impact. As a result, some of the acceleration measured on the valve plate will be coherent with the
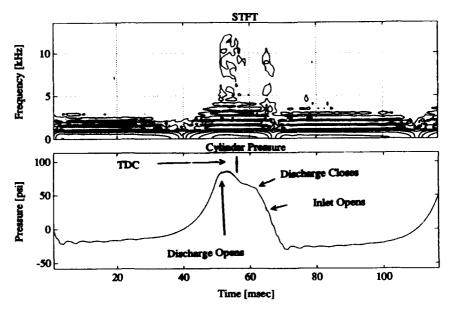
355

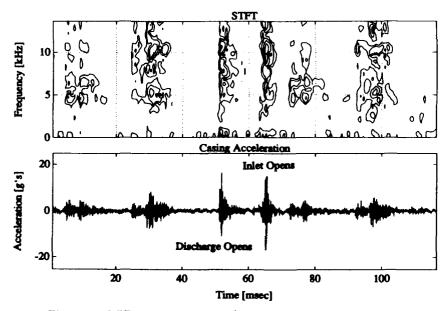Figure 6: STFT and time-waveform of the cylinder pressure.

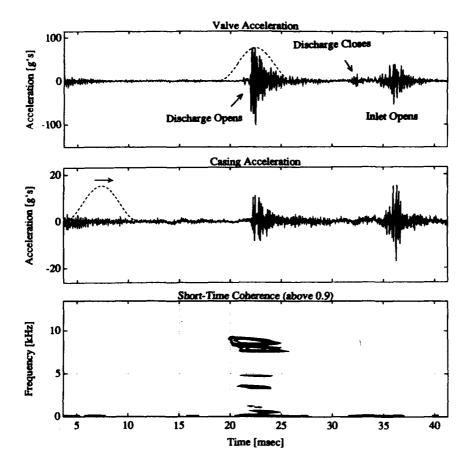Figure 7: STFT and time-waveform of the casing acceleration.

Figure 8: A. Valve body acceleration with analysis window; B. Casing acceleration with first analysis window; C. Short-time coherence (above 0.9).

casing acceleration simply because they are both responses to other sources within the compressor. We must avoid regions of high coherence with sources other than the one of interest.

To obtain the STC we start by windowing out the event of interest in one signal, like the valve body acceleration due to valve impact shown in Figure 8a. To the other signal, like the casing acceleration shown in Figure 8b, we apply a "sliding" window such as that used to obtain the STFT. We have used Hanning windows here. The STC is then simply the ordinary coherence function evaluated between the single windowed source event and each of the windowed sections of the output signal. Figure 8c shows a contour plot of the STC between the valve acceleration and casing acceleration for levels of coherence 0.9 and above in the frequency range from DC to 13.2 kHz. The results shown are for40 independent time records obtained over consecutive machine cycles. The STC shows that there are several regions of high coherence, each lasting for about 5 msec, the largest existing between approximately 7 kHz and 9 kHz.

357

In order to assure ourselves that the high coherence was not due to some source other than the valve impacts, we generated STCs between the valve and casing accelerations for other vibration events, such as the opening of the adjacent inlet valve. The vibration caused by the inlet valve impact, labeled in the accelerations shown in Figures 5 and 7, is significant in both the valve plate and the casing. The STC between these two signals for this event is highest below 2 kHz but then falls off rapidly, becoming insignificant above 4 kHz. We have also examined the STC between the valve body acceleration and the pressure waveform discontinuities to locate regions of high coherence between these two source events, and found that they are strongly coherent below 2 kHz, with a rapid falloff in coherence above 2 kHz. Thus assured, we have begun pursuing the recovery of valve impact forces in the 7 to 9 kHz range and are finding encouraging results.

Conclusions and Future Work: The impulsive waveform recovery technique outlined and demonstrated in this paper shows promise for use as a diagnostics tool for reciprocating machinery, especially when coupled with the short-time coherence for determining the joint time-frequency ranges over which meaningful recoveries can be made. In the near future we expect to successfully complete our investigation into the recovery of the valve impact waveforms, and will then work on recovery of an impulsive waveform related to the pressure discontinuities in the cylinders and manifolds. Once this work is complete we will begin to introduce known, but non-destructive, faults into the compressor to see how well they can be detected using this technique.

References

1. R.H. Lyon, *Machinery Noise and Diagnostics*, Butterworth, Boston, 1986.

2. A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989, pp. 209-211.

3. D.L. Bowen and R. H. Lyon, "Recovery of Diagnostic Signals for Machinery Transients," Paper presented at the International Machinery Monitoring and Diagnostic Conference, September 11-14, 1989, Las Vegas, Nevada.

4. M. Tohyama, R.H. Lyon, and T. Koike, "Pulse Waveform Recovery in a Reverberant Condition," J. Acoust. Soc. Am., 91(5), pp. 2805-2812, May 1992.

5. Oppenheim, p. 784.

6. Oppenheim, p. 807.

7. J.S. Bendat and A.G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd Edition, Wiley, NY, 1986.

REAL-TIME VIBRATION HEALTH MONITORING
FOR TRANSIENTLY OPERATING HIGH-SPEED TURBOMACHINERY

Sid W. Hite, III
Sverdrup Technology, Inc./AEDC Group
1103 Avenue B, Arnold AFB, TN 37389-1400

Abstract: The Department of Defense (DoD) has placed increased emphasis on innovation and prototyping of cutting-edge weapons systems. Hence, more of the engines undergoing altitude testing at the Arnold Engineering Development Center (AEDC) Engine Test Facility (ETF) will fall into the high-cost, high-risk classification.

Successes with after-the-fact diagnosis of failure modes in jet engines led to the initiation of a project to identify potential hardware problems **before** catastrophic failures occur. The vibration-based **Health Monitoring System** (HEMOS) is an expert system which will continuously monitor vibration signatures for symptoms of component faults. The system will further allow trending capability during user-defined data "windows."

This paper reviews the findings of the literature/technology survey, details established system requirements, describes the proposed operating system, and relates analytical results from conducted research.

Key Words: Amplitude; Analysis; Expert System; Faults; Frequency; Health Monitoring; Instrumentation; Transient; Vibration

Introduction

Background: In this day of shrinking defense budgets, the Department of Defense (DoD) has placed increased emphasis on innovation and prototyping of cutting-edge weapons systems. The philosophy dictates incorporation of advanced technologies into systems capable of doing more ... with less. Current research and development efforts in the aerospace propulsion arena aim at doubling engine thrust while cutting specific fuel consumption (SFC) in half. These lofty goals are imbedded in the Increased Performance Turbine Engine Technology (IPTET) program. In order to meet the IPTET goals, manufacturers are turning to new materials, manufacturing processes, and cycle optimization techniques. Advancing the state-of-the-art necessitates construction of prototypes which are both costly and high-risk.

The Engine Test Facility at Arnold Engineering Development Center (AEDC), Arnold AFB, TN, is dedicated to altitude testing of aerospace propulsion systems in support of prototyping, demonstration, development, qualification, initial flight release, flight test, and component improvement. With primary DoD emphasis on prototyping and demonstration, a higher percentage of test articles at the AEDC will fall into the high-

---

cost, high-risk classification. Successes with post-mortem fault diagnoses through vibration analysis have prompted AEDC to seek a means of real-time fault identification. The goal: prevent catastrophic failures of multimillion dollar engines.

Problem Statement: The principles of predictive maintenance have long been applied to rotating machinery in the paper, power, and chemical industries. Such machinery usually operates at steady-state conditions for long periods. Hence, developing machine component faults tend to appear as changes to the vibratory response characteristic of the machine. The task of deciphering developing faults through vibration monitoring becomes much more difficult, however, when transiently operating high-speed turbomachines (such as aircraft engines) are involved. Aircraft engines, by nature, are extremely transient machines. Requirements to operate over a wide range of altitudes and flight velocities translate into an extensive matrix of inlet conditions to the machine (pressure, temperature, density, etc.). Since vibratory responses may vary considerably with one or more of these factors, a huge array of data may be required to define a "baseline" vibration signature for a specific engine model. The problem is further complicated by the fact that it is difficult to implement a system which can digitally sample the analog sensor outputs fast enough to accurately describe the "true" signature when operating conditions are rapidly changing.

Many other problems associated with vibration monitoring during transient operation have been identified and investigated. Three separate activities have led AEDC to determine that the problems associated with such a system are surmountable. First, post-mortem failure analyses identified indications of component faults minutes and hours, respectively, prior to catastrophic failures of turbomachines at AEDC. The demonstrated ability to predict the failure modes prior to teardown inspections led to much-needed support for this project. Second, a literature survey and feasibility study revealed promising work which strives to circumvent the pitfalls associated with monitoring transiently operating turbomachinery. Third, significant advances in digital sampling of analog signals have been made at AEDC through the use of massively paralleled processing techniques.

This paper will review early studies which indicated an AEDC Health Monitoring System (HEMOS) was indeed feasible, discuss system requirements definition, describe the data processing vehicle which may allow fruition of the HEMOS goal, and convey analytical results which aided the AEDC focus toward developing engine health criteria and a prototype HEMOS.

Literature and Technology Survey: The survey included review of approximately forty articles, papers, and documents and five different online or offline operating systems. Objectives of the survey included: (1) investigating current systems capable of performing online or offline data acquisition, reduction, analysis, and/or diagnosis; (2) determining the overall benefits of such systems; (3) evaluating the highlights and limitations of these systems; and (4) investigating methodologies previously employed to ascertain machine health through monitoring of performance and vibration parameters.

Premier work accomplished in the realm of aircraft engines has been done by the Royal Air Force (RAF) and Rolls Royce (RR) on the Adour and RB199 fighter engines. Because of its relevance to an AEDC monitoring system, the RAF/RR system bears discussion in some detail. This work focused on reducing vibration test time and teardown/rebuild necessitated after overhaul. Outstanding results have been achieved as the RAF/RR system is capable of identifying the predominant faults on these engines in near real-time.

The RAF/RR system relies on an enhanced database of vibration data. Initially, accel/decel data were taped for 83 different engines. Of these, nine engines had significant hardware faults. The "healthy engine" criteria were derived from the 74 engines found to be without mechanical faults. "Unhealthy engine" criteria were originally derived from the nine faulty engines, and the database is continually updated with field/overhaul fault occurrences and results from engines with induced faults.

Pattern matching with so-called "fault curves" has been instrumental in achieving a factor of seven reduction in vibration test time and a drastic reduction in mechanical failures at the engine and component levels (Carr, 1990).

Although the AEDC goal of preventing catastrophic failures of test articles is slightly different from the RAF/RR goals of decreasing test time and spare parts costs, the AEDC HEMOS system will closely parallel the RAF/RR methodologies in terms of identifying and diagnosing faults.

The technology survey included attending demonstrations of several "in-place" automated vibration-based monitoring systems and conducting personal interviews with many experts in the fields of vibration data acquisition, reduction, and analysis. Operating systems reviewed include: (1.) Bentley Nevada System 64 in use by the AEDC Facility Operations and Maintenance organization; (2.) Strain Gage Monitoring System (SGMS) developed for the Compressor Research Facility (CRF) at Wright Laboratories by Mechanical Technologies, Inc. (MTI); (3.) Automated Vibration Diagnostics (AViD) system developed by MTI and used in vibration acceptance testing at the Oklahoma City Air Logistics Center (OCALC); and (4.) a prototype Computer Assisted Dynamic Data Monitoring and Analysis System (CADDMAS). Roundtable discussions and personal interviews were conducted with representatives of Wright Labs, IRD Mechanalysis, General Electric Aircraft Engines, MTI, University of Tennessee - Knoxville, CSI, OCALC, and AEDC.

Although a lengthy discussion is beyond the scope of this paper, the technology survey dramatically increased the AEDC understanding of the problems associated with automated monitoring systems as they apply to transiently operating turbomachinery. Further, many of the discussions with experts in the vibration field revealed potential solutions to these problems, thus significantly influencing the definition of system requirements for the AEDC HEMOS system.

Requirements Definition for a Prototype HEMOS System

System Overview: A schematic overview of the planned HEMOS system is shown in Fig. 1. The system should first Fast Fourier Transform (FFT) accelerometer, velocimeter, and/or proximity probe data, and then time synchronous merge this digitized analog data with predetermined transient digital data parameters. These transient parameters may include various pressures, temperatures, speeds, flight conditions, and variable geometry positions. The system will be able to operate in both continuous **and trend** modes.

In the **continuous** mode, acquired vibration and transient data are continually merged and passed to the host computer system, where health monitoring algorithms may be applied to the processed data. Currently, algorithms are intended to check vibrations versus manufacturer's specified limits and screen for potential rotor dynamic, gear box, and bearing faults. If no potential problems are identified, the merged data remain in a circular file to be overwritten. Should a potential problem be identified, however, an alarm system will identify the channel(s) in an overlimit condition, and the data from all channels will be written to
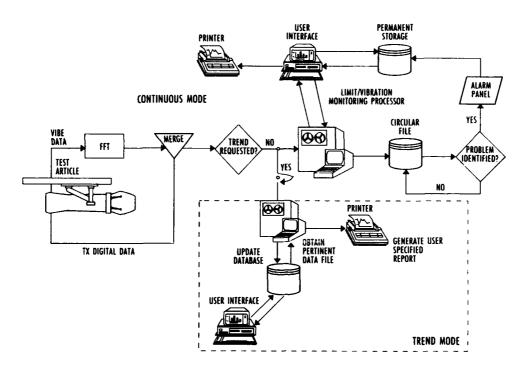
Fig. 1. **Health Monitoring System** (HEMOS) overview.

a file for permanent storage. In the case of an alarm condition, the circular file data will also be dumped immediately to permanent storage to provide a 20-min history of the engine conditions leading up to a fault. A user interface will be provided to: (1) input necessary information for HEMOS algorithms, and (2) allow interaction for user interrogation of the permanent storage file. Continuous mode will be operational whenever the engine is rotating, and HEMOS should be capable of performing all monitoring and alarm functions within 1 sec of data acquisition.

The **trend** mode of HEMOS operation is to be invoked upon user demand to provide a historical data trending capability during defined data "windows." Potential windows may include engine starts/shutdowns, 2 min accels and decels at health check flight conditions, baseline vibration data at military power, etc. The trend data software will compute statistical variations of current data with the historical database and generate a user-specified hardcopy comparison or CRT display for the vibration analyst(s). Trend mode must allow visibility of pertinent trend information for all channels upon user demand within 3 min of data capture without interrupting the flow of data through the continuous mode algorithms. A user interface will once again be necessary to specify input files, output format, and data window start/stop times. This interface may be the same for both modes.

Data Validity Checking: The first step following data acquisition is to apply some method of checking for erroneous data. Electronic noise has long been an enemy to the vibrations analyst, and left unchecked false alarms could completely undermine user confidence in any automated health monitoring system.

Rocketdyne's Automated Dynamic Data Analysis and Management System (ADDAM) is an integrated acquisition, digitization, mass storage, and offline analysis system which incorporates a subroutine for identification of obvious noise sources. The sources include electronic line noise (i.e., 60 Hz), cable whip, and broadband "white noise" (Tarn, 1987). Mr. Barney Bare of MTI and Dr. Belle Upadhaya of UT-Knoxville recommend synchronous averaging to limit noise in the spectra (personal interviews, 1991). The AEDC-developed CADDMAS currently employs a comparison of discrete frequency levels versus overall root-mean-square (rms) voltage to detect obviously erroneous data. The AEDC HEMOS system will incorporate some combination of the above to minimize false alarm indications due to spurious data.

Input Channels: The HEMOS system will be capable of accepting and processing FFTs for a total of 12 dynamic data channels with frequency range to 8 KHz, including a worst-case scenario of 12 accelerometers in 12 different sensor locations. Figure 2 illustrates a typical altitude test array of turbine engine vibration sensors. HEMOS must also accept and process a key-phasor signal from each rotor shaft for phase relation determination in diagnosing shaft cracks and for component balancing exercises.
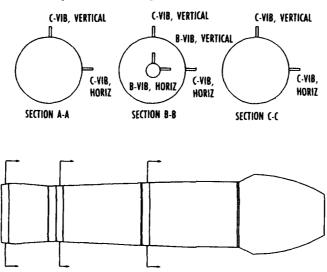
Fig. 2. Typical vibration sensor array for an altitude test
article. (Bearing sensors may/may not be located in
each plane show.

Further, the system should accept up to 20 transient digital data input channels at sample rates to 1,000 sps. Time synchronous merging of the digitized analog data and the transient digital data is to be accomplished based upon IRIG time within the data resolution of the digital data.

Physical Quantities - Acceleration, Velocity, or Displacement: White makes a strong case for defining limits in terms of peak velocity. He argues that velocity is directly proportional to the energy of vibration and is independent of frequency in the energy equation (White, 1970). AEDC experience, however, indicates acceleration measurements are necessary at frequencies above ≈ 1 KHz, because acceleration responses indicative of rolling element bearing faults, gear faults, and airfoil resonances are "in the mud" of the velocity amplitude resolution.

Additionally, some AEDC test articles are equipped with various proximity sensors, and the HEMOS system must be able to accept them.

Therefore, the AEDC HEMOS will accept dynamic data inputs in terms of acceleration, velocity, and/or displacement. It is also required that integration capability exist such that data measured in terms of acceleration may be processed to velocity and displacement units or that velocity data may be integrated to produce displacement data. No requirement currently exists for differentiation of the waveform signal as significant errors are generally introduced during this process.

Frequency and Amplitude Resolution Requirements: Dynamic data frequency response should range from 0 to 8 KHz with ± 5 Hz resolution. Parabolic interpolation will be employed to enhance resolution. Any phase distortion introduced by the HEMOS hardware must be quantified and corrected in all data presentations.

HEMOS processing functions will resolve the vibration amplitudes to within ± 5 percent of the maximum peak regardless of the units employed. Similarly, computation and application of imposed vibration limits should allow specification of amplitude limits in acceleration, velocity, and/or displacement units.

Processing of dynamic data will include user-specified windowing functions to include Max Flat Top, Hanning, and Rectangular windows, among others.

Data Storage Requirements: The circular data file containing "non-event" data in the continuous mode should store up to 20 min of data from 12 dynamic and 20 transient data channels (maximum of 1.7GB required) before overwrite begins. The capability should exist for user-demanded download to permanent storage. The permanent storage file associated with the continuous mode of operation should accept and store up to 6 hr of data per 14 air-on-hour test for 12 dynamic and 20 transient data channels (maximum of 30.1GB required). This file is to be accessible by the health monitoring algorithms when an "event" is identified and also upon user demand through the interface. File manipulation capability through user interaction is also required.

The historical database associated with the trend mode should have enough storage capacity to maintain vibration histories of Overall, 1/rev NL, and 1/rev NH vibration levels versus speed and/or time for six different data windows (i.e., health check points of 2- to 3-min duration each) for all dynamic data parameters installed on a given test article for the length of the current test program. Additionally, a baseline vibration signature (Overall and 1/rev fan and core responses) must be maintained for a given sensor location on a specific engine model during those six different data windows. It is anticipated that no more than 2 hr of trend data will be acquired and processed for each dynamic data channel during a test program (≈ 100 engine hr). This requirement necessitates an additional 7.1GB of storage capacity. Therefore, total storage capability for the circular file, continuous mode data, and historical trend data is 38.9GB.

Identifiable Machine Faults: HEMOS will focus on rotor dynamic-related faults. Shaft faults which should be identifiable include a bent or bowed shaft (usually 1X low rotor or high rotor speeds, NL or NH) and coupling slop (presence of 1/2X). Module faults include tip rubs of fan, compressor, and turbine blades and may be identified by a multiple of NL or NH frequency which corresponds to the number of blades on the rotating stage. Assembled rotor faults include out-of-balance conditions induced by improper bearing loads and diagnosed

through changes in critical speeds or vibratory responses (between consecutive accels, for instance).

HEMOS will be capable of computing and screening for bearing fault frequencies indicative of pending failures at the subcomponent level. Bearing misalignment (excessive 2X NL or NH) and oil whip/whirl (subsynchronous NL or NH) are still other faults which may be identified by spectral screening. Gearbox faults may include tooth defects (periodic spikes at gear mesh frequency) and eccentricity of the gear mesh.

Vibration Sensor Placement: The HEMOS system must be adaptable to the instrumentation configurations chosen by the AEDC users. Obviously, a case-mounted accelerometer may not be sensitive to internal bearing faults, but may easily sense rotor out-of-balance, bow, or bent shaft conditions. Access to algorithms for deducing bearing and gear faults will be limited to data from sensors in proximity to the bearing housings and gear boxes. The goal of the HEMOS will be adaptability from engine family to engine family and configuration to configuration.

Limit Application Capability: Due to the transient nature of aircraft engines, a methodology must be developed for comparison of vibratory responses to established limits. To investigate further, consider a "typical" turbofan engine with a low rotor operating speed regime of 3,000 to 6,000 rpm and refer to Fig. 3.

At 3,000 rpm (Fig. 3a, top);
      assume a 1X response of 2 mils p-p at 50 Hz (3,000 rpm)
      and a 2X response of 1 mil p-p at 100 Hz.
This response characteristic may be indicative of a relatively rough running rotor (2 mils 1X at idle) with bearings which are poorly aligned to the shaft.

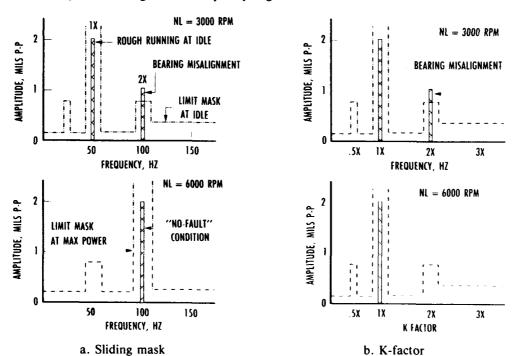a. Sliding mask   b. K-factor

Fig. 3. Limit application methodology.

Now, accelerate to 6,000 rpm (Figure 3a, bottom):
    a 1X response of 2 mils p-p occurs at 100 Hz (6,000 rpm)
    and no significant 2X component of vibration is noted.
This response characteristic represents a healthy engine which is operating well within the vibration limits of most manufacturers.

In the first case, a 100-Hz response of 1 mil p-p was interpreted as a fault (bearing misalignment), while in the second case a 100-Hz response of 2 mils p-p was deemed to be "normal." Consequently, the limit application methodology in a turbine engine vibration monitoring system must be able to recognize the various responses in the spectrum and apply the appropriate limit.

This suggests a "Sliding Mask" limit application technique. Theoretically, a different limit mask would exist for every combination of low and high rotor speeds, and an expert system would be required to continuously identify the significant spectral responses and apply the appropriate limits. Note that the limit mask must slide to the right in the frequency domain as the engine is accelerated from 3,000 rpm (Fig. 3a, top) to 6,000 rpm (Fig. 3a, bottom).

By employing the "K-Factor" approach to limit application illustrated in Fig. 3b, AEDC seeks to greatly simplify this problem. The "K-Factor" approach draws on the fact that all rotor dynamic responses are related to rotor speed in an integral or nonintegral manner, such that
    freq = K-Factor * N/60
    or,
    K-Factor = freq * 60/N
where N = high or low rotor speed and
    K-Factor = constant related to geometry or phenomena
    freq = frequency in Hz
For integral vibrations, K-Factor is simply an integer multiple of engine speed. For nonintegral vibrations, K-Factor is a mixed fractional number (i.e., K-Factor $\approx$ 0.47 for vibration due to oil whirl phenomenon). These responses are easily computed, so if the nominal response range for a given engine family can be characterized in terms of Amplitude versus K-Factor, then HEMOS may be programmed to interrogate for potential problems using a single limit mask and avoid the huge development task associated with a "Sliding Mask."

Data Presentation Alternatives: The HEMOS system will be capable of providing all usual vibration data presentation formats including (but not limited to): spectra, trending plots, engine order tracking plots, waterfall plots, Campbell diagrams, orbits, Bode' plot, Nyquist plots, tables, alarm synopses, etc.

CADDMAS - A Vehicle for HEMOS

CADDMAS Overview: The Computer Assisted Dynamic Data Monitoring and Analysis System (CADDMAS) is an AEDC system which is being developed to acquire, store, process, and display dynamic signals from engines under evaluation in the AEDC Engine Test Facility. CADDMAS was first envisioned to tackle the huge task of processing strain-gage aeromechanical data and displaying results in near real-time. Huge volumes of data are currently generated at high sample rates (i.e., signal analysis to 32 KHz), CADDMAS uses a network of smart Integrated Sensors for preprocessing and a parallel architecture for additional processing and display to deliver engineering diagrams online and on demand from the user.

CADDMAS is designed to provide the computational horsepower to accomplish online visibility of analog parameters which drive test direction and ensure test article hardware health through component monitoring. A highly successful prototype CADDMAS was demonstrated in 1992 and is being used for test support in the ETF. The prototype system consists of dynamic data processing capabilities for 12 data channels sampled to provide analysis to 20 KHz. The system has been used to produce thousands of Campbell Diagrams, spectral envelopes, and tracking plots in an on-line fashion with delivery to the end user in mere seconds. Similar off-line processing techniques may take up to two weeks to produce the same quality and quantity of information.

The AEDC Directorate of Technology - Propulsion Division (DOTP) end product CADDMAS will be capable of acquiring, processing, and displaying 48 channels to 50 KHz and an additional 24 channels to 20 KHz. The system will further be able to accept 32 transient digital data parameters at rates up to 1,000 sps.

CADDMAS is defining a new state-of-the-art for real-time dynamic data processing and analysis. With its astounding computational power, the system has many potential uses beyond online test monitoring of aeromechanical data. In fact, the current prototypical capabilities provide a stable vehicle on which to base the HEMOS system. Table 1 overviews the HEMOS system requirements and the corresponding CADDMAS capabilities.

Table 1. CADDMAS Capabilities Versus HEMOS Requirements

	CADDMAS CAPABILITIES	HEMOS REQUIREMENTS
DATA VALIDITY CHECK	FREQUENCY VERSUS RMS	REQUIRED
NO. INPUT CHANNELS	48 DYNAMIC	12 DYNAMIC
	40 DIGITAL	20 DIGITAL
PHYSICAL QUANTITIES	ACCELERATION VELOCITY DISPLAY	ACCELERATION VELOCITY DISPLAY
FREQUENCY RANGE	0-50 KHz	0-8 KHz
FREQUENCY RESOLUTION	±20 Hz	±5KHz
AMPLITUDE RESOLUTION	±2 PERCENT	±5 PERCENT
DATA STORAGE	70+GB	38.9GB
LIMIT APPLICATION	TBD	K-FACTOR APPROACH
PLOT ALTERNATIVES	TBD	MANY

TBD INDICATES: TO BE DEVELOPED

HEMOS Development: The ETF dynamic data acquisition, processing, production, and analysis community stays abreast of current work through a Dynamic Data Working Group. Early in the HEMOS feasibility study, it was recognized that CADDMAS might be an ideal candidate to provide the data processing capability for a vibration-based health monitoring system. Hence, HEMOS and CADDMAS personnel have maintained close contact throughout the requirements definition phase. In FY93, hardware and software development for a prototype HEMOS system has begun as a subtask of the CADDMAS project under DOTP. The forthcoming section will review results from early vibration studies conducted at AEDC.

HEMOS - Analytical Results and Focus for the Future

Background: Before an expert system may be programmed to identify abnormal conditions based upon vibratory spectra, "normal" vibratory responses must first be characterized. Analysis was conducted using reduced vibration data from a typical air-breathing turbofan engine. The goals of this analysis were to identify the engine operating parameters which have a primary or secondary effect on the vibratory characteristics of various engine components.

In order to limit the scope of effort, data from two accelerometers were reduced and analyzed. One accelerometer was internally mounted on the housing of the high rotor shaft thrust bearing while the other was case-mounted at the engine front frame in the vicinity of the fan. These vertically oriented sensors were chosen because they were the most responsive accelerometers to internal and external vibrations, respectively. For simplicity, we will designate the bearing-mounted accelerometer B-VIB and the case-mounted accelerometer C-VIB. For the purposes of this paper, primary consideration will be given to the bearing vibration analysis.

Effects of Engine Operating Parameters on Bearing Vibrations: Preliminary investigation showed that several engine operating parameters influence the vibratory response characteristics measured by the bearing accelerometer. The primary response measured was always the 1/rev signal generated by the residual unbalance of the high rotor system. This is expected, since the B-VIB was mounted on the axial thrust bearing of the high rotor system. The function of this bearing is to restrain forward thrust - transmitting unbalance energy out of the engine through the frame struts in the form of vibration. Vibration amplitudes measured by B-VIB appear to decrease with increasing inlet pressure (Fig. 4), and the 1/rev response increases with increasing inlet temperature (Fig. 5). Further investigation into these trends, however, yields an important result.

At each of the two higher inlet pressure conditions shown in Fig. 4, the engine is operating at a control- specified pressure limit, and fan rotor speed has been rolled back to maintain engine operation at or below this limit. Since the fan rotor and core rotor are aerodynamically coupled, this results in a lower core speed as well. The result is a lower vibration amplitude measured at the bearing housing, because the residual mass unbalance is rotating at a lower speed at higher inlet pressures (for identical power settings).

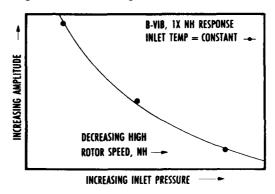
Fig. 4. Effects of inlet pressure on bearing maximum vibratory repsonse.
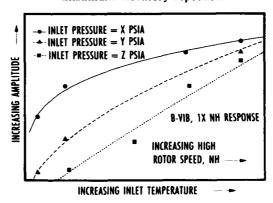
Fig. 5. Effects of inlet temperature on bearing maximum vibratory response.

368

Data trends in Fig. 5 indicate increasing vibratory amplitude with increasing inlet temperature for three different inlet pressures. Once again, these trends are actually related to core speed. The fan speed schedule for most turbofan engine families is primarily a function of inlet temperature subject to various pressure, temperature, and speed limitations. Fan speed increases with increasing temperature (until limits are incurred) aerodynamically driving core speed higher as well.

Similar trends hold for lube oil pressure and temperature (data not shown). Vibration amplitude increases with increasing lube pressure, but further investigation reveals that this trend is also related to speed. The lube pump is driven by the core shaft through the power take-off (PTO) shaft and gearbox. Consequently, higher core speeds result in higher pump speeds and higher lube tank pressures. The vibrations once again increase with increasing core speed. Likewise, lube temperature has only a secondary effect on trends of vibration amplitude. Increasing vibration with increasing lube temperature is again related to core speed through the gearbox.

Although several parameters were found to have a secondary effect on vibrations measured by B-VIB, the primary effect is always due to core rotor speed. In the absence of operation at a critical speed (which are generally designed to be outside the engine operating regime), the highest vibratory amplitudes may be expected at the highest speeds and may be attributed to residual mass unbalance in the rotor. This is a significant result, because it greatly simplifies the approach necessary to adequately monitor bearing health.

If the "normal" range of vibratory amplitudes can be identified for each family of engines, then it should be possible to screen for abnormalities based on 1/rev vibration and its harmonics. Addition of a capability to calculate and screen for the bearing fault frequencies will supplement the 1/rev monitoring, and a bearing health monitoring scheme will thus be implemented via the HEMOS bearing algorithm.

Effects of Engine Operating Parameters on Case Vibrations: Measured front frame vibratory responses react primarily to mass unbalance of the low rotor (1/rev NL), acoustic (dynamic pressure) excitations, and wake shedding. Similar results are expected for case-mounted sensors along the length of the engine, with responses occurring based upon the proximity of the accelerometer to major excitation sources (i.e., 1/rev NL or NH, blade passing, augmenter rumble or screech, etc.).

Engine manufacturers have well-developed limits for 1/rev NL and NH, and incorporation of these limits into the HEMOS methodology will be simple. Although more complicated, expected resonant crossings for various components due to acoustic or wake shedding excitation may be computed. Parametric studies will be conducted to determine the range of response magnitudes attributable to such resonances.

For example, vibratory stresses in front frame struts naturally induce a vibratory response measured at the case by C-VIB. If HEMOS is programmed to expect these resonances and associated increase in vibrations, false alarms will be kept to a minimum. Again, the difficulty lies in characterizing the expected range of amplitudes for each resonant response.

Analysis Results - Response Amplitude Repeatability: Representative plots of the variation in response amplitude versus frequency for the B-VIB and C-VIB accelerometers are included as Fig. 6. For consecutive decels at like inlet conditions, B-VIB variation ranged from 0 to 8 percent for measured responses above the noise floor. Response variation measured at the

369

front frame by C-VIB ranged from 0 to 22 percent during the same consecutive decels. Similar variations were noted at other flight conditions as well.

An investigation of accel/decel response amplitude variation was also conducted. Although the data are not included, variations ranged from 0 to 22 percent for B-VIB and from 3 to 10 percent for C-VIB, respectively.

Due to significant differences in bearing loads between accel and decel operation in some military engine families, further analysis will have to be completed before meaningful results may be gleaned for incorporation into HEMOS.

Analysis Results- Necessity for Automation: The analysis results reported herein were a significant undertaking. Limiting the effort to two data channels for a minimal number of engine data acquisition events allowed certain trends and conclusions to be drawn, but much is yet to be learned. Couple this level of effort with the fact that output from only two accelerometers at 21 flight conditions (accels/decels at each) was analyzed, and one begins to see the enormity of analysis required to characterize the vibration responses over the flight map.
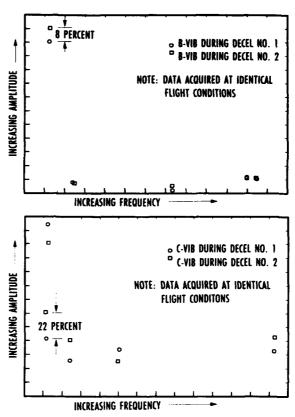
Fig. 6. Response amplitude repeatability versus frequency for bearing (top) and case)bottom) accelerometers.

The necessity for automating the analysis process becomes apparent when we realize that many test articles are delivered to AEDC with up to 12 accelerometers, and many of the test programs encompass 50 to 60 flight conditions. The HEMOS algorithms will initially be developed and applied to offline data. A database of expected vibratory responses will be acquired for each type of engine on test at AEDC. Capabili:ies will be imbedded which statistically characterize the range of Amplitude versus K-factor data so that amplitude limits may be assigned.

Go Forward Plan: Initial work focused on determining the feasibility of developing an automated, vibration-based expert system for monitoring the health of transiently operating turbomachines. More recently, requirements for the HEMOS system have been developed and initial analytical studies completed. The focus for the future includes:
    1. developing and encoding the HEMOS algorithms;
    2. adapting the logic and algorithms into a functional prototype capable of analyzing, condensing, and characterizing vibration health data; and

370

3. applying the HEMOS prototype to taped vibration data to begin characterization of expected vibratory responses for the engine families tested at AEDC.

Summary and Conclusions

Unlike the paper, power, and chemical industries, where predictive maintenance payoff is largely in avoiding lost production, the goal of the AEDC Health Monitoring System is to avoid catastrophic failures of multimillion dollar jet engines. Correct post-mortem fault diagnoses through analysis of vibration data were the impetus for investigating such a system, and much work has ensued.

First, a summary of the literature/technology survey shows that systems do exist which circumvent problems associated with real-time health monitoring. Most notably, the Royal Air Force and Rolls Royce have enjoyed remarkable success in reduction of vibration test time and spare parts by applying an expert system to diagnose frequent component faults.

Second, HEMOS requirements have been specified in terms of data acquisition, processing, resolution, storage, and monitoring. Additional key issues which have been addressed include sensor placement, interrogation "windows" for trending, and limit application methodology.

Third, it appears that by building on the capabilities of CADDMAS hardware and software, there will be no need to purchase sophisticated off-the-shelf hardware to provide the HEMOS skeleton. The vibration health monitoring function will be incororated as a facet of CADDMAS capabilities.

Finally, analysis results were presented which evaluate the influence of engine operating parameters on bearing and case vibrations. It was determined that the primary influnce on bearing vibrations is rotor speed. In the absence of critical speed operation, maximum bearing vibrations may be expected at the highest speeds. This is a significant result which suggests that the "normal" bearing vibrations are much less a function of flight condition than they are of rotor speed. Thus, the approach necessary to adequately monitor bearing health has been greatly simplified.

Measured front frame vibratory responses were found to react primarily to mass unbalance of the low rotor (1/rev NL) and to acoustic excitations of the front frame. This suggests a need to incorporate inlet conditions into the HEMOS logic, since acoustically driven mechanical resonances of the engine frames may be predicted if the temperatures, pressures, and modal frequencies are known. The analysis effort has highlighted the need for an automated technique, and current efforts aim at developing this capability.

Bibliography

Carr, H. R. "UK Military Engine Vibration Monitoring Experience." SAE E-32 Engine Health Monitoring Committee Symposium—Vibration Monitoring and Analysis, Zurich, Switzerland, October 23-24, 1990.

Garcia, R. C., Jr. "An Expert System to Analyze High Frequency Dependent Data for the Space Shuttle Main Engine Turbopumps." NASA-CP-2492 Part 1, Third Annual Conference on Artificial Intelligence for Space Applications, November 2-3, 1987.

Harris, W. J., Minnear, J., Chang, Dr. J. D., and Ziebarth, H. K. "Airborne Engine Condition Analysis Instrumentation." Naval Air Systems Command Final Report AD-904725L REPT-72-8518, Garrett Airesearch Manufacturing Company, July 1972.

Lagace, L. and Hayes, P. "Computer-Assisted Strain-Gage Monitoring and Data Reduction System." WRDC-TR-90-2064 (F33615-85-C-2503), Final Report, August 1990.

Tarn, R. B. "Rocketdyne Automated Dynamics Data Analysis and Management System." NASA-CP-2488 Vol. 2, NASA 58th Shock and Vibration Symposium, Huntsville, AL, October 13-15, 1987, pp. 151-152.

White, R. F. "Vibratory Velocity; A Measure of the Destructive Potential of Vibration." Allison Laboratory Report, A-9628, August 1970, pp. 15-18.

A FUZZY LOGIC AND NEURAL NETWORK SYSTEM FOR BENEFIT ANALYSIS OF COST REDUCTION IMPROVEMENTS

S. Nils Straatveit
General Dynamics/Electric Boat
Groton, Connecticut 06340

Abstract: Budget reduction in the military and competitiveness and profit margins in the civilian sector are putting pressure on maintenance functions to reduce costs. Solutions, such as changes to the design, operation or maintenance procedures that will reduce labor and material cost are available. However, even though engineering instincts may be correct, up front costs and vaguely supportable benefits are often not well received by management. The benefit analysis problem is further compounded by effects from interrelated processes and parameter uncertainty. This paper defines a multi-criterion decision-making methodology which accounts for uncertainty by utilizing fuzzy logic and automates the analysis by utilizing neural network technologies. Neural Networks also provide the ability to accomplish model free estimation of the complex interactions in the systems under study. The analysis methodology further supports linguistic as well as numeric input and will provide an audit trail to enable management support of the cost benefit improvements.
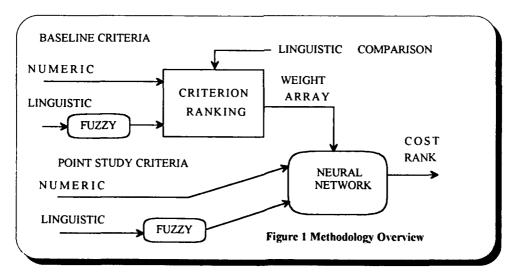
Introduction: Management needs supportable analysis for decisions to purchase new equipment, modify processes or utilize new technology. An engineered analysis with an audit trail, focused on cost evaluation, is needed to assess improvements reputed to reduce life cycle cost. Some methodologies that have been put forth are very subjective and thereby lose the ability to carry the point. These methods also suffer because of their mismatch with the problem morphology of multi-criterion decision-making with uncertainty. Furthermore, the problem criteria are not often easily specified in numerical form or common units; rather they are better specified verbally. To date then, subjective approaches are somewhat matched but lack in their ability to effectively communicate or provide minimal auditing capability and are difficult to update as new aspects of the problem are learned. Numerical automation approaches are ill fitted to the problem

morphology of linguistic input parameters for multi-criterion decision-making with uncertainty.

This paper defines a methodology that supports multi-criterion decision-making with uncertainty and satisfies the requirements from management. Specifically, many of the inputs for cost benefit analysis are linguistic variables and therefore would be better represented as fuzzy variables. Automation of the analysis is also desirable so that, as discussions develop new criteria can be added to the neural network structure and decision makers can quickly get another analytical run with an audit trail supporting the process.

Methodology: The evaluation methodology requires the collection and processing of relevant criteria with the objective of focusing all the parameters to a relative cost comparison; see Figure 1. The methodology utilizes fuzzy logic and neural network technologies to automate the input and processing of the criteria values. Criteria are selected and hierarchically organized by field experts. The structure of the neural network used reflects the criteria hierarchical organization. Linguistic criteria input is supported using fuzzy logic. A ranking process, also using fuzzy logic, establishes the initial weights and dependency rules for the assessment. The organized criteria and weights confirm the baseline by being a reflection of the present system characteristics. The system under study can include maintenance procedures as well as components such as pumps or motors. Changes, to be analyzed, can cause the addition of criteria, but are weighted for a null impact on the baseline system. During each system analysis or point study the evaluation process will provide an audit trail in the form of criterion decision weights so that management can support benefit implementations.
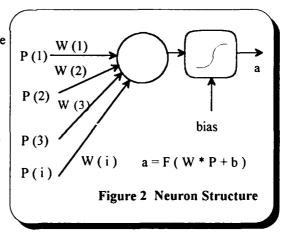
Technology: The use of fuzzy logic and neural networks as systems estimators that do not require an initial model is of high value to the complex problem described above. The network used is a fuzzy version of the multi-layer perceptron. Fuzzy logic is also used to

Figure 1 Methodology Overview

allow linguistic input for both the criteria value input and control of the initial criteria ranking process. As the initial weights are established and implemented, in the neural network, learning is accomplished using the gradient-descent-based back-propagation learning algorithm. Then a dynamic programming algorithm is used to compare and verify the learning process and, most importantly, to assess weights at each layer for support of the audit trail information.

Neural networks: A basic definition of neural networks comes from W. Newman [1990]. "Neural networks are a class of algorithms that can be modeled as an array of fairly simple interconnected circuits called "neurons," much like the neuron interconnections of the nervous system in the brain. Just as with the brain, a neural network can be configured to be trainable. In other words, a neural network can compare its output under controlled conditions with a desired signal and adjust various internal weights to minimize the differences between the actual output and the training signal." In addition neural networks process data fast and efficiently due to their massively parallel construction. Neural networks can also represent nonlinear systems without the user specifying a mathematical model (model free estimators), can process incomplete information, and are robust in applications due to their fault tolerance. These attributes contribute to fulfillment of the problem requirements above.
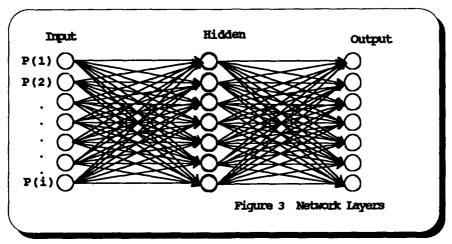
Neural networks are composed of neurons arranged in layers. Each neuron is structured, as shown in Figure 2, with weighted inputs being summed and then output through a function usually a sigmoid. The sigmoid function provides the ability to represent nonlinear processes and to provide a continuous valued output. A bias input is also included with the weighted inputs to improve stability when the network is in the learning mode.

$$a = F (W * P + b)$$

Figure 2 Neuron Structure

A typical network might use three layers, an input layer- the data presentation layer, a hidden layer, and an output layer. The optimal number of hidden layers and the number of neurons in each layer is mostly empirical. In the configuration described, neurons operate in parallel in a layer and are found in the hidden and output layer. All neurons in one layer connect to each neuron in the next layer as shown in Figure 3.

Learning is accomplished in the network using the gradient-descent back-propagation learning algorithm. Rumelhart and Hinton [1986] created the algorithm by generalizing the Widrow-Hoff learning rule, a gradient descent procedure, to multiple layer

375

Figure 3 Network Layers

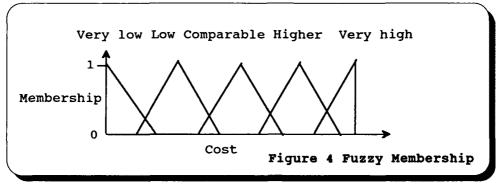networks and nonlinear differentiable transfer functions. Gradient descent continually changes the values of the network weights and biases in the direction of steepest descent with respect to error. Changes in weight and bias are proportional to that neuron's effect on the sum squared error of the network.

Fuzzy logic: Fuzzy logic is in broad and effective use in Japan to instantiate complex functions for products that have limited processing capability. Fuzzy logic is effective because it is the "logic of interpolative reasoning" (Zedeh). Interpolation is achieved by using class-of-membership functions, fuzzy inferencing, and a host of defuzzification methods. According to Zadeh, interpolation can reduce the solution of a large system to a series of equations that can be arrived at linguistically and whose multiple concurrent solutions are interpolated and defuzzified to arrive at a single answer. These techniques allow engineers to design systems that implement satisfactory, approximate answers to large system problems with much shorter design cycles than conventional methods.

Concepts from fuzzy sets are incorporated at various stages in the methodology and in the creation of the fuzzy version of the neural network.. Input data handling of the criterion can be done as exact (numerical) and/or inexact (linguistic) forms using the fuzzy neural network input process specified by S. K. Pal and S. Mitra [1992]. Fuzzy sets model uncertain or ambiguous data so often encountered in real life and simplify the processioning of complex interactions consisting of imprecise or incomplete information. In such cases it may become convenient to use linguistic variables and hedges such as low, medium, high, very, and more or less to augment or even replace numerical input information.

The components of the input vector consist of the membership values to the overlapping partitions of linguistic properties low, medium, and high corresponding to each input criteria. Certain domains may require the use of a five-term set such as {very small, small, medium, large, and very large}; see Figure 4. This approach provides the scope for incorporating linguistic information and increases robustness in tackling imprecise or

Figure 4 Fuzzy Membership

uncertain input specifications. In Pal and Mitra, once the membership values for the criterion have been computed, the actual numerical values are no longer needed or used. The benefit of this approach is utilized, in our methodology, between groupings of related criteria. However, we maintain the numerical values also in order to follow the input criterion construction methodology of Paek and Lee [1992]. Here the data of each basic criterion are represented in numeric fashion in order to develop initial weights for the neural network to use. Neural networks learn faster if initial weights can be established. The process to determine the initial weights uses the analytical hierarchy process developed by Saaty [1988]. A matrix is formed to compare criterion i with criterion j. Experts in the field make decisions of relative cost on a pair-wise basis for the selected criterion. Fuzzy logic as a linguistic input device is used in the criterion ranking. Linguistic parameters such as much lower, lower, comparable, higher, much higher are used to communicate the relative costs between pairs of criteria for a particular point study.
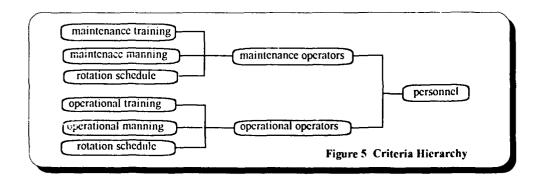
Initial weights are used to initialize the neural network. The neural network will be organized in architecture to follow the organizational hierarchy of the evaluation criteria. The fore mentioned back-progation learning algorithm is well known and used here to further tune the weights in a straight forward manner. Learning speed is always a concern and improvements are available. Presently, the Karhunen-Loe've Transformation may be used as in Malki and Moghaddamjoo [1991]. In this approach, an initial set of training vectors is obtained by applying the transformation on the training data. The training is started in the direction of the major eigenvectors of the correlation matrix of training patterns and then continues by gradually including the remaining components, in their order of significance. However, processes in the hidden layers need to be observed for audit purposes. The dynamic programming approach to optimal weight selection is used to continue the process. The rationale is that "a multi-layer feed-forward neural network can be thought of as a multistage decision process since optimal selection of weights for each layer is akin to an optimal choice of decisions at each stage, and weights in a layer can affect only the outputs of subsequent layers, as with decisions in a multistage decision process" Saratchandran [1991]. By virtue of this process, the impact of the various criteria, collected at each stage, on the final output can be recorded for the audit trail.

Improvements or changes to the system will be generated, applied to the network and examined with respect to the plausibility of the output conclusion. For example, a change proposed would have varying ranges of medium to high development costs. This should lead to a negative or lessor choice of selection for this proposal as compared to the baseline system. Having verified the methodology performance by this process, more complex modifications representing real historical changes will be processed.

Example: To begin the process a systems engineer would evaluate the technical attributes of systems that are large cost drivers in life cycle maintenance. Criteria that might impact maintenance cost are:

- operational availability
- operational manning
- maintenance manning
- safety
- essentiality
- failure severity
- repairability
- maintainability
- redundancy
- information timeliness
- information accuracy
- repair induced failures
- parts costs
- operational training
- maintenance training
- logistics tail
- parts commonality
- failure detection

After the evaluation criteria are established they are organized hierarchically; see Figure 5. The architecture of the fuzzy neural network will follow this form.

Figure 5 Criteria Hierarchy

378

In the investigation it may be found that scheduled maintenance is causing components with no degradation to be removed and shipped to a depot for repairs. An improvement to the maintenance process then, as expressed in a paper by Cieri and Elfont [1991], would be to collect data or monitor the operational item to establish its condition and need for repair before removal. This changes the maintenance process from scheduled repair to a potentially cost saving one of Reliability Centered Maintenance. With this potential cost savings in mind, the system, including components and procedures, is analyzed to identify relevant criteria.

Now one needs to evaluate the cost impact of developing and deploying a monitoring system. Criteria, such as development cost, need to be added. More detailed criteria such as the diagnostic accuracy of the monitoring system should also be added in order to evaluate competing monitoring system proposals. Even the monitoring systems life cycle cost should be included. In the proposed methodology these criteria are added to the analysis with a null contribution to the cost for the baseline system. The evaluation then proceeds with the comparison of the various monitoring system concepts and the baseline. Ranking will indicate the benefit, if it exists, of using monitoring systems and which of the potential monitoring system choices is more cost effective and why.

Conclusion: A methodology for benefit analysis of equipment life cycle cost reduction improvements has been defined. The methodology utilizes fuzzy logic and neural network technology. The methodology accepts numerical and linguistic input for criteria using fuzzy logic membership functions. Neural network technology is used to characterize the system, which includes procedures and hardware components, and provides a cost ranking output.

The benefit of this approach to the analysis of improvements for cost reduction in maintenance methodologies, processes, and equipment is in its ability to match the problems characteristics of multi-criterion decision making with uncertainty, and provide management with an audit trail to support their decisions.

Due to time and space limitations there are problem details not included in this paper. Such as, the inclinations of the membership value functions in the fuzzy logic for correct direction of criteria impact, the process of learning the index functions to compensate for different units in the criteria, the present value calculations of future savings or future costs and the time value of money, the combined use of fuzzy logic and neural processing (especially the class-of-membership functions and fuzzy rule generation to utilize the linguistic solution contribution to the evaluations). As always a certain degree of cut-and-try hand optimization is usually required, not only of the membership functions but of the scaling between the physical variables and the input and output variables of the fuzzy system.

References:

1. A. M. Cieri and M. M. Elfont, "Engineered Approach to Effective Maintenance Management", Naval Engineers Journal, May 1991.

2. J. H. Paek and Y. W. Lee, "Selection of Design/Build Proposal Using Fuzzy-Logic System", Journal of Construction Engineering and Management, Vol. 118, No. 2, June 1992.

3. S. K. Pal and S. Mitra, "Multilayer Perceptron, Fuzzy Sets, and Classification", IEEE Transactions on Neural Networks, Vol. 3, No. 5, September 1992.

4. W. C. Newman, "Detecting Speech With an Adaptive Neural Network", Electronic Design March 22, 1990.

5. P. Saratchandran, "Dynamic Programming Approach to Optimal Weight Selection in Multilayer Neural Networks", IEEE Transactions on Neural Networks, Vol. 2, No. 4, July 1991.

6. T. L. Saaty, "Multicriteria Decision Making: The Analytic Hierarchy Process", University of Pittsburgh Press, Pittsburgh, Pa., 1988.

7. D. E. Rumelhart and G. E. Hinton, "Learning Internal Representations By Error Propagation," D. Rumelhart and J. McCelland, editors, Parallel Data Processing, Vol. 1, Chapter 8, the M.I.T. Press, Cambridge, MA 1986.

8. L. A. Zadeh, "Fuzzy Sets," Information and Control, 8(3), pages 338-353, 1965.

9. H. A. Malki and A. Moghaddamjoo, "Using the Karhunen-Loe've Transformation in the Back-Propagation Training Algorithm", IEEE Transactions on Neural Networks, Vol. 2, No. 1, January 1991.

APPENDIX

MFPG PUBLICATIONS

Both printed and microfiche copies of the following MFPG publications whose catalog numbers start with either **AD** or **COM** may be obtained from the

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161

Glossary of Terms	AD 721 354
Proceedings of Meeting Nos. 1-9 (set of five)	AD 721 359

Meeting Nos. 1-5
  Papers and Discussion on Failure Analysis and Control

Meeting No 6 "Detection, Diagnosis and Prognosis"
      December 6, 1968

Meeting No 7 "Failure Mechanisms as Identified with Helicopter Transmissions"
      March 27, 1969

Meeting No 8 "Critical Failure Problem Areas in the Aircraft Gas Turbine Engine"
      June 25-26, 1969

Meeting No 9 "Potential for Reduction of Mechanical FailureThrough Design Methodology"
      November 5-6, 1969

Proceedings

Meeting No 10	"Vibration Analysis Systems" January 21-22, 1970	AD 721 912
Meeting No 11	"Failure Mechanisms: Fatigue" April 7-8, 1970	AD 724 475
Meeting No 12	"Identification and Prevention of Mechanical Failures in Internal Combustion Engines" July 8-9, 1970	AD 721 913
Meeting No 13	"Standards as a Design Tool in Surface Specification for Mechanical Components and Structures" October 19-20, 1970	AD 724 637
Meeting No 14	"Advances in Decision-Making Processes in Detection, Diagnosis and Prognosis" January 25-26, 1971	AD 721 355
Meeting No 15	"Failure Mechanisms: Corrosion" April 14-15, 1971	AD 725 200
Meeting No 16	"Mechanical Failure Prevention Through Lubricating Oil Analysis" November 2-4, 1971	AD 738 855

Meeting No 17 "Effects of Environment Upon Mechanical Failures, Mechanisms
    and Detection"   AD 750 411
    April 25-27, 1972

Meeting No 18 "Detection, Diagnosis and Prognosis"   AD 772 082
    November 8-10, 1972

Meeting No 19 "The Role of Cavitation in Mechanical Failures" (NBS SP 394)
    October 31-November 2, 1973   COM-74-50523

Printed copies of the following publications may be obtained from the

<div align="center">

U.S. Government Printing Office
Superintendent of Documents
Washington, DC 20402

</div>

Microfiche copies of these publications may be obtained from the NTIS.

Proceedings

Meeting No 20 "Mechanical Failure - Definition of the Problem" (NBS SP 423)
    May 8-10, 1974   SN003-003-01451-6

Meeting No 21 "Success by Design: Progress Through Failure Analysis" (NBS SP 433)
    November 7-8, 1974   SN003-003-01639-0

Meeting No 22 "Detection, Diagnosis and Prognosis" (NBS SP 436)   SN003-003-01556-3
    April 23-25, 1975

Meeting No 23 "The Role of Coatings in the Prevention of Mechanical Failure" (NBS SP 452)
    October 29-31, 1975   SN003-003-01664-1

Meeting No 24 "Prevention of Failures in Coal Conversion Systems" (NBS SP 468)
    April 21-24, 1976   SN003-003-01760-4

Meeting No 25 "Engineering Design" (NBS SP 487)   SN003-003-01829-5
    November 3-5, 1976

Meeting No 26 "Detection, Diagnosis and Prognosis" (NBS SP 494)   SN003-003-01844-9
    May 17-19, 1977
Meeting No 27 "Product Durability and Life" (NBS SP 514)   SN003-003-01935-6
    November 1-3, 1977

Meeting No 28 "Detection, Diagnosis and Prognosis" (NBS SP 547)   SN003-003-02083-4
    November 28-30, 1978

Meeting No 29 "Advanced Composites" (NBS SP 563)   SN003-003-02120-2
    May 23-25, 1979

Meeting No 30 "Joint Conference on Measurements and Standards for
    Recycled Oil/SystemsPerformance and Durability" (NBS SP 584)
    October 23-26, 1979   SN003-003-02272-1

Meeting No 31 "Failure Prevention in Ground Transportation Systems" (NBS SP 621)
April 22-24, 1980   SN003-003-02428-7

Meeting No 32 "Detection, Diagnosis and Prognosis: Contribution to the
Energy Challenge" (NBS SP 622)   SN003-003-02361-2
October 7-9, 1980

Meeting No 33 "Innovation for Maintenance Technology Improvements" (NBS SP 640)
April 21-23, 1981   SN003-003-02425-2

Meeting No 34 "Damage Prevention in the Transportation Environment" (NBS SP 652)
October 21-23, 1981   SN003-003-02488-1

Printed copies of the following MFPG publications are available from

Cambridge University Press
110 Midland Avenue
Port Chester, NY 10573

Proceedings

Meeting No 35 "Time Dependent Failure Mechanisms and Assessment Methodologies"
April 20-22,1982

Meeting No 36 "Technology Advances in Engineering and Their Impact on Detection,
Diagnosis and Prognosis Methods"
December 6-10, 1982

Meeting No 37 "Mechanical Properties, Performance and Failure Modes of Coatings"
May 10-12, 1983

Meeting No 38 The Proceedings was not published because of the format of the meeting.

Meeting No 39 "Failure Mechanisms in High Performance Materials"
May 1-3, 1984

Meeting No 40 "Use of New Technology to Improve Mechanical Readiness, Reliability and
Maintainability"
April 16-18, 1985

Meeting No 41 "Detection, Diagnosis and Prognosis of Rotating Machinery to Improve
Reliability, Maintainability, and Readiness Through the Application of
New and Innovative Techniques"
October 28-30, 1986

Meeting No 42 The Proceedings was not published. Inquire at the Vibration Institute
regarding the purchase of copies of individual papers.

Meeting No. 43 "Advanced Technology in Failure Prevention"
October 3-6, 1988

Printed copies of the following MFPG publications are available from the

Vibration Institute
6262 S. Kingery Hwy
Suite 212
Willowbrook, IL 60514

Proceedings

Meeting No 44 "Current Practices and Trends in Mechanical Failure Prevention"
April 3-5, 1990

Meeting No. 45 "Focus on Mechanical Failures: Mechanisms and Detection"
April 9-11, 1991

Meeting No. 46 "Economic Implications of Mechanical Failure Prevention"
April 7-9, 1992

Meeting No. 47 "The Systems Engineering Approach to Mechanical Failure Prevention"
April 13-15, 1993

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS None	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT UNCLASSIFIED/UNLIMITED	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S) MFPG 47		5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A	
6a. NAME OF PERFORMING ORGANIZATION Vibration Institute	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research	
6c. ADDRESS (City, State, and ZIP Code) 6262 S. Kingery Highway Willowbrook, IL 60514		7b. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000		10. SOURCE OF FUNDING NUMBERS	

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

The Systems Engineering Approach to Mechanical Failure Prevention

12. PERSONAL AUTHOR(S)
Compiled by Henry C. Pusey and Sallie C. Pusey

13a. TYPE OF REPORT Proceedings	13b. TIME COVERED FROM 93/1/1 TO 93/4/15	14. DATE OF REPORT (Year, Month, Day) 93/4/15	15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Rotating equipment, reciprocating machinery, motors, expert systems, failure detection, diagnostics, failure analysis, failure prediction, failure testing, signature analysis

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This document is the proceedings of the 47th Meeting of the Mechanical Failures Prevention Group (MFPG) which was held in Virginia Beach, VA on April 13-15, 1993. The proceedings contains featured papers from the Opening Session and a Plenary Session. There are session papers on concurrent engineering for mechanical systems, nondestructive evaluation and information processing, failure mechanisms and life extension, and condition based maintenance systems engineering.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Peter Schmidt	22b. TELEPHONE (Include Area Code) (202) 696-4410	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted
All other editions are obsolete   SECURITY CLASSIFICATION OF THIS PAGE

☆ U.S. Government Printing Office: 1986—607-044